White matter hyperintensities and cognition in Parkinson's disease

Masters Thesis

Author: Sophie Halliday

Supervisors: Dr Steve Marsh & Dr Tracy Melzer

Master of Science in Medical Physics,

Department of Physics and Astronomy,

The University of Canterbury

White matter hyperintensities and cognition in Parkinson's disease

Sophie Halliday

Abstract

Cognitive decline and dementia are common non-motor impairments associated with Parkinson's disease (PD). White matter hyperintensities (WMHs) are vascular MRI findings that are common in older individuals and have been shown to be associated with cognitive impairment. This study aims to assess the predictive power of WMH volume and spatial location on cognitive impairments in PD using an automated WMH identification algorithm and MRI.

A training data set of 40 subjects with high WMH volume were used to assess the performance of 4 automated WMH detection algorithms against gold-standard manually identified WMH maps. The optimal algorithm was then applied to a longitudinal PD cohort (PD=207, HC=51) and cognitive impairment and dementia prediction were investigated using Bayesian regression modelling and model comparisons. Specifically, I investigated the relationship between global and regional WMH volumes, global cognitive ability, and individual cognitive domain scores.

Brain Intensity AbNormality Classification Algorithm (BIANCA) was the identification algorithm that resulted in the most accurate and precise WMH maps in our training set. BIANCA was used with optimised parameters to extract WMH maps for the whole cohort. I found no evidence of increased global WMH burden in PD relative to healthy controls, nor any difference across cognitive ability within PD. Furthermore, I found no significant spatial relationship between WMH volume and global cognition, however anterior periventricular WMHs (PVWMHs) associated with attention and executive function cognitive domains.

These results suggest that WMHs are not a clinically-relevant biomarker of cognitive impairment in PD. While there is a weak emerging correlation between increased anterior PVWMH volume

and specific cognitive domain dysfunction, resounding support of regional WMHs prediction of cognitive domain dysfuncytion is lacking.

Acknowledgements

My gratitude goes to my supervisors; Dr Tracy Melzer and Dr Steve Marsh who have provided me with guidance and direction over the past year. Their positivity and support has been outstanding and has made working on this thesis most enjoyable.

Thanks to the entire team at New Zealand Brain Research Institute for the support offered to me throughout this project. Special thanks to the phenomenal team members had a hand in this project including collecting cognitive data and acquiring medical images; this work would not have been possible without them. A special thanks to Dr Campbell Le Heron and Dr Reza Shoorangiz and their clinical and computing expertise.

Lastly, to my family and friends, for the love and care you have given me during this year in particular, my endless thanks goes to them.

Contents

List of Figures

List of Tables

Glossary

BIANCA = Brain Intensity AbNormality Classification Algorithm

BRM = Bayesian Regression Model

DARTEL = Diffeomorphic Anatomical Registration Through Exponentiated Lie

DER = Detection Error Rate

FPR = False Positive Ratio

FNR = False Negative Ratio

FSL = FMRIB Software Library

GLM = General Linear Model

KOLDIC = k-Fold Information Criteria

LGA = Lesion Growth Algorithm

LOO = Leave-one-out

LPA = Lesion Probability Algorithm

LPM = Lesion Probability Map

MNI = Montreal Neurological Institute

MRI = Magnetic Resonance Imaging

PD = Parkinson's Disease

PD-MCI = Parkinson's Disease with Mild Cognitive Impairment

PD-N = Parkinson's Disease with Normal Cognition

PDD = Parkinson's Disease Dementia

PVWMH = Periventricular White Matter Hyperintensities

SD = Standard Deviation

SI = Similarity Index

SPM = Statistical Parametric Mapping

TFCE = Threshold-Free Cluster Enhancement

TPM = Tissue Probability Map

UBO = Unidentified Bright Object

WMH = White Matter Hyperintensities

Chapter 1

Introduction

1.1 Parkinson's disease

1.1.1 Overview

Parkinson's disease (PD) is the second most common neurodegenerative disease affecting the population, with an estimated 1 in 500 individuals in New Zealand with PD in 2013 [33]. PD is a debilitating disease, and its prevalence increases exponentially with age [26], which is of particular importance in a community with an aging population. Historically, PD has been viewed as a motor disorder, however, development in the understanding of PD in recent decades indicates a myriad of non-motor impairments are also associated with the disease [34].

1.1.2 Pathology

PD is classically characterised by hallmark motor symptoms; namely resting tremor, bradykinesia, gait impairment, and rigidity [26]. The associated non-motor symptoms that have emerged in recent decades include, but are not limited to, olfactory impairment, sleep and cognitive impairments, hallucinations, depression, apathy, anxiety, and autonomic dysfunction [26, 35]. PD is understood to be a multi-system disorder, clinically heterogeneous in nature and associated with a number of protein aggregates and neuromodulatory systems involved [26].

The basal ganglia (BG) are subcortical structures deep within the forebrain that are made up of a

number of nuclei that each play a different role across a number of motor and cognitive processes [17]. The neurological pathways of the BG contribute to the execution or inhibition of movement via projections to the frontal cortex through the thalamic nuclei [17]. The substantia nigra (SN), so-called due to its blackened appearance compared to surrounding tissues, is a nucleus in the BG and is separated into two regions; the dopamine-rich pars compacta (SNpc), and the pars reticular (SNpr). See Figure 1.1 for the basic structure of BG and constituents. Dopaminergic neurons are housed within the SNpc and project to the caudate-putamen, with dopamine playing a vital role in voluntary motor movement, mood, and emotion control [9, 17]. A reduction in the quantity of dopamine produced is thus closely related to the disruption of the systems underlying these functions. In PD, mass cell degeneration and death of dopamine neurons in the SNpc contribute to the depletion of dopamine that occurs with the development of the disease and it is estimated that an individual with PD will lose 50-80% dopaminergic cells in the duration of the disease [5]. The mechanism of dopaminergic neuron death in PD remains unknown, resulting in treatment strategies focused on alleviating symptoms, rather than altering the underlying cause of the disease.

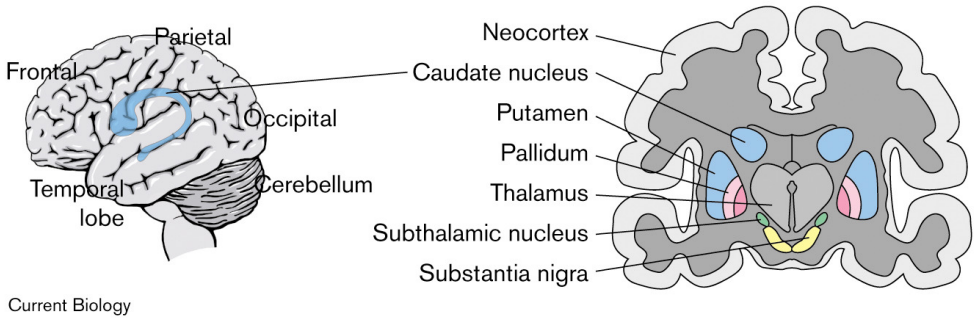


Figure 1.1: Basic basal ganglia and nuclei structure. Striatum (blue), pallidum (pink), subthalamic nuclei (green), and substantia nigra (yellow) [17]

The current gold standard for diagnosis of PD is in postmortem investigation of the presence of misfolded α-synuclein protein and degradation of the SNpc [26]. In practice, the diagnosis is made clinically at times supported by adjunctive methods such as DAT-SPECT imaging, with correct clinical diagnosis of PD pathology in 84% of cases [52, 60]. Although the development in imaging techniques in past decades have provided the potential for in vivo disease identification, there are currently no unequivocally recognised biomarkers of PD that can be identified in neuroimaging

techniques such as MRI [29].

To date, treatment avenues are focused on symptomatic regulation as opposed to disease cure [26]. Treatment focus is placed on improving a patient's quality of life and controlling the inhibiting symptoms when possible. The areas of disease manifestation that are focused on for treatment include the motor symptoms, psychosis, mood, and autonomic dysfunction [10]. Pharmaceutical treatment options that target the issue of dopamine deficiency are largely used to treat the motor symptoms of PD, with common pharmaceutical options including Levodopa and dopamine agonists [9, 10, 26]. It is common that numerous symptoms require treatment as heterogeneity of symptoms and progression make a single treatment an unlikely solution for this disease.

1.1.3 Cognition in Parkinson's disease

Cognitive impairment is a non-motor problem that affects a large number of PD patients and it is estimated that 80% of PD patients will develop dementia over the course of their disease [5, 21]. One framework for considering cognitive impairment in PD groups patients into those with normal cognition (PD-N), those with mild cognitive impairment (PD-MCI), and those with dementia (PDD). Assessment an individual's overall cognitive status often requires rigorous cognitive testing to assesses all cognitive domains, including executive function, attention, visuospatial, learning and memory, and language domains [1]. There are international guidelines from the Movement Disorder Society (MDS) task group for the diagnosis of the cognitive status of PD patients. PD-MCI refers to an individual with PD who experiences minimal impairment in daily life but whose cognitive test scores in at least one cognitive domain fall below 1.5 SD of the mean age-adjusted normative data or control data [13]. PD-MCI is considered an intermediate cognitive stage between normal cognition and dementia. PDD refers to a PD patient who experiences significant interruption to everyday activities due to cognitive impairments, not caused by motor impairments. Clinically, PDD is defined by cognitive scores 2 SD below the mean of age-adjusted normative data in at least two cognitive domains. Many studies have characterised the mechanisms of cognitive decline in PD including both postmortem and in vivo studies [1], however, no singular definitive cause of cognitive decline in PD has been determined yet.

One theory of the characterisation of cognitive decline in PD is the dual syndrome hypothesis. This characterisation details degradation of specific cortical regions resulting in specific cognitive domain impairments. The hypothesis differentiates between two independent PD syndromes; the first that fronto-striatal dysfunction is indicative of planning, working memory, and executive function impairments, and the second, that non-frontal dysfunction is indicative of visuospatial impairment [27, 48]. In addition to visuospatial impairments, it is suggested that non-frontal syndrome is thought to lead from PD-MCI to PDD more rapidly than the frontal syndrome [22]. The validity of this hypothesis is of keen interest in the characterisation of cognitive impairment in PD and will be investigated in this thesis.

Neuropsychological testing is insufficient to identify those individuals most at-risk of progressing quickly from PD-MCI to PDD, thus additional information is required if we are to identify the most at risk patients. Identification of this group with heightened PDD development risk is advantageous as novel treatments or trials can be targeted towards the group. Neuroimaging is a powerful and promising technique which may provide us with this additional information required to determine the state of an individual's brain health at present and in the future. Neuroimaging also has the potential to be a tool for identifying biomarkers of cognitive impairment in PD. Such biomarkers of cognitive decline and dementia are urgently needed to enrich treatment samples with those more likely to develop dementia in the near future, as well as potentially to assess treatment effectiveness. White matter hyperintensities are one such biomarker holding PD cognition prediction potential.

1.2 White matter hyperintensities

1.2.1 Overview

Cerebral small vessel disease (CSVD) is an umbrella term to describe brain pathologies that are through to arise from changes within cerebral arterioles, capillaries, and small blood vessels [49]. They are visible on MR images with manifestations including, but not limited to, subcortical infarcts (lacunae), cerebral microbleeds, and white matter hyperintensities (WMHs) [49]. Developments in medical imaging techniques have improved our ability to visualise CSVD and the associated expressions of the pathology, allowing for an investigation into their cause, development, and consequences.

WMHs appear as bright areas of white matter (WM) visible on MR images, specifically, proton density weighted, T2-weighted, and T2-weighted fluid-attenuated inversion recovery (FLAIR) images. The properties of these MR techniques make them sensitive to vascular damage [8, 37]. WMHs have been observed in individuals of varying levels of health, but they are consistently associated with increasing age, and traditional cardiovascular risk factors such as hypertension [24, 38, 53], and the presence of WMHs has been related to impairments in balance, mobility, and cognition in otherwise healthy individuals [5, 49]. In addition to these reported impairments, it has been suggested that the presence of WMHs modulate the severity of a number of diseases like stroke, Multiple Sclerosis (MS), Alzheimer's disease (AD), and PD [24].

WMHs can be present through the entire mass of white matter, however, they tend to follow a characteristic distribution. There are regions of WM that are generally not affected (e.g. juxtacortical) but regions around the ventricles are often observed [36]. Total WMH burden can thus be separated into deep WMHs (DWMHs) and periventricular WMHs (PVWMHs) dependant on their distance from the ventricles. While the definition of WMHs spatial location has been developed, questions remain on the relative importance of these WMH classifications in terms of pathologies and impairments [18].

1.2.2 White matter hyperintensity identification

For WMH volume to be used in any research or clinical capacity, accurate identification is of utmost importance. Historically, WMH burden has been assessed by an experienced grader using various scales used to classify WMHs; Fazekas scale [15] and Scheltens scale [45] for example. These scales are reliable WMH identification methods when followed by an experienced grader, however, they do not provide any information on the specific location of WMH burden in an individual [19].

Technological developments in previous years have made manual segmentation of WMHs an option for burden assessment in which a grader can manually identify WMH voxels on a medical image, slice by slice. This technique has the advantage of providing spatial information and accurate qualification about the WMHs that the previously mentioned rating scales cannot provide. While

such manual rating techniques have high success in the identification of WMHs, they are time-consuming and can be variable based on inter- and intra- grader differences [19, 46]. Furthermore, large cohorts and medical images of high spatial resolution render manual identification impractical because of the time-consuming nature of the work. Therefore, there is a need for a detection technique that is immune to the issues faced by manual grading while also being reliable and robust.

There are a number of automated and semi-automated algorithms available for WMH identification, yet there is no gold standard identification method, much less one specific for WMHs in PD. The available literature for these automated identification techniques reports success in their respective cohorts, however, variability in data such as MR acquisition, image quality, and cohort can all contribute to the variability of success when any given technique is used on another cohort [19, 24, 47, 51]. The lack of a gold-standard automatic identification technique also means that different papers use different identification techniques which limits comparability, which further contributes to the confusion surrounding the significance of WMHs on specific pathologies.

1.2.3 White matter hyperintensities and cognition

In MS, WMHs have been shown to be related to "irreversible neurological disability" by damage to WM tracts that connect the cortex to nuclei (such as the constituents of the BG) [4], and WM lesions are a known biomarker of the disease [47]. There are many automated and semi-automated WMH identification algorithms that have been created specifically for WM lesion identification in MS. WMHs have also been shown to exasperate cognitive decline in AD, with this relationship being widely accepted and agreed upon [8, 16]. In addition to contributing to worsening cognition in cohorts with neurological disease, there is also an association between healthy individuals' worsening cognition and increased WMH burden [28].

The association between WMHs and cognitive impairment and decline in PD is based on the effect WMHs have on cortical connection and WMH tracts that have been found in other diseases, such as MS as mentioned above. Studies have been carried out claiming that WMHs may cause or worsen cognition in PD with PVWMHs potentially interrupting ascending thalamocortical and descending corticospinal fibres causing cognitive impairments [5]. Spatial location of WMHs, therefore, may

have different clinical consequences corresponding to the association fibres in the region of WM damage [50].

The dual syndrome hypothesis, as introduced above, suggests that spatial location of WMHs could be an important influencer of different cognitive dysfunction patterns. In terms of WMHs, this hypothesis suggests that PVWMH burden could be indicative of specific cognitive impairments in PD, allowing for categorisation of the cognitive impairment. Specifically, it is suggested that anterior PVWMHs cause executive function and attention domain dysfunction, while posterior PVWMHs cause memory and visuospatial domain dysfunction.

A recent study by Reginold and colleagues investigated the diffusion properties of WM tracts in the presence of WMHs and reported that WM fibre tracts that crossed WMHs exhibited abnormal diffusion characteristics, which can be related back to the dual syndrome hypothesis with WMH location associated with dysfunctional WM fibres [41]. The study also found that diffusion characteristics of WM tracts worsened when they were located close to a WMH. These findings indicate there is potential benefit in knowing the location of WMHs in an individual to specify their particular PD related cognitive impairments, future cognitive decline, and assist in identifying patients that can be targeted for novel treatments as they are developed. This being said, research into the relationship between WMHs and spatial location and the combined effect on cognition in PD is still needed.

The literature reviewed for this study indicated a divide in findings of correlation between WMHs and cognitive impairment in PD. While some studies reported significant correlation between WMHs and cognitive impairments, others fail to emulate the same findings [12, 16, 31, 53]. For this reason, an investigation into WMHs in PD in our large, longitudinal PD cohort is of interest due to the potential predictive power of WMHs of cognition in PD.

1.3 Study rationale and breakdown

The aim of this thesis was to investigate the importance of WMHs as a biomarker of cognitive impairment in PD. MRI and automated WMH identification algorithms were used to detect WMHs

in a large PD cohort. First, in a training cohort, I tested four WMH identification algorithms and compared them to 'gold standard' manually segmented WMHs map in order to select the optimal automated method. I then applied the optimal method to the full cohort, producing WMHs maps in all individuals. Next, I investigated the relationship between WMHs, PD, and cognitive impairment, in a cross-sectional and longitudinal manner.

The motivation for determining the relationship between WMHs and cognitive impairment is the potential of finding a biomarker that indicates an individual's risk of developing cognitive impairment in PD. With the research into the correlation of cognition and WMH in PD unequivocal to date, this study aims to detail the correlation of WMH volume to cognition in our large PD cohort. In addition, an objective marker of disease severity (i.e. WMHs) could also be used to assess the effectiveness of any novel treatments.

This thesis is set out in 2 parts:

Part 1: Training, assessment, and selection of a method to automatically identify WMHs within our cohort.

Part 2: Application of the automated algorithm identified in Part 1 to all participants to determine:

- Do WMHs differ between PD and controls cross-sectionally?
- Do WMHs differ across cognitive groups (i.e. control, PD-N, PD-MCI, PDD) at both the global and regional level?
- Does global and/or regional WMH volume can predict future global cognition?
- Does global and/or regional WMH volume can predict future cognitive domain scores?

Chapter 2

Magnetic resonance imaging

An overview of magnetic resonance imaging (MRI) is presented here as the imaging modality is utilised extensively in this study. MRI an imagining technique that utilises magnetic properties of protons to create highly detailed diagnostic images. MRI exploits the presence of hydrogen in the body, with differences in hydrogen content in various tissues contributing to image contrast. A hydrogen nucleus has an associated magnetic moment produced by the nuclear spin of the constituent proton about its axis. When subject to external magnetic field changes, the magnetic moments exhibit certain properties which can be perturbed and recorded to yield an MRI signal.

A sample of magnetic moments will self arrange in a disorganised, random orientation in which all magnetic moments cancel each other out, resulting in no net magnetisation. However, when exposed to an external magnetic field, B_0, the protons tend to align with the external magnetic field, yielding a net magnetisation, M along the direction of B_0. In addition, the external magnetic field exerts a torque on the protons that causes an alteration in spin precession proportional to B_0 and the gyromagnetic ratio of the particular element, $\gamma = 42.58MHzT^{-1}$ for hydrogen [7]. The angular frequency of this spin precession is calculated by the Lamour equation and is a cornerstone in MRI (Equation 2.1)

$$\omega = \gamma B_0 \tag{2.1}$$

.

The Lamour frequency ω is also the resonant frequency of the sample. Exposure to an external radiofrequency (RF) pulse at the resonant (Lamour) frequency causes the net longitudinal magnetisation to rotate away from the longitudinal (along the z-axis) into the transverse plane (the xy-plane). The flip angle is the angle between the original net magnetisation and the net magnetisation vector when the resonant RF pulse is turned off. The net magnetisation in the xy-plane continues to precess and induces a current in a receiver coil. This orientation is not sustainable, so the protons will eventually revert to the original equilibrium and the strong signal from the rotating net longitudinal mangetisation precessing in the xy-plane decays. Precession of the net magnetisation in the transverse plane (xy-plane) produces the raw MR signal. The properties of the sample will determine the rate of the return to equilibrium, which is recorded and used in the MR image construction [7].

2.0.1 Relaxation

Relaxation is the process by which an excited sample loses energy and returns to equilibrium; there are two chief methods of relaxation in MRI.

- **T2 relaxation** or **spin-spin relaxation** in which energy obtained by a sample from the resonant RF pulse is dissipated through inhomogeneities in the sample.
- **T1 relaxation** or **spin-lattice** in which energy obtained by a sample from the resonant RF pulse is released into the lattice of surrounding molecules.

Factors dictating both T1 and T2 relaxation times include molecule size, motion, and interactions. Figure 2.1 indicates how the two relaxation times change dependant on these factors, and a consistently longer T1 relaxation time is observed. Another fundamental principle behind MRI is that different tissue types exhibit different characteristic T1 and T2 relaxation times. MRI can, therefore, be tuned to accentuate these differences to create images exhibiting different contrasts and encompassing different biological information, discussed below in Section 2.0.2.2.

2.0.2 MRI acquisition parameters and pulse sequences

There are a number of core parameters that can be manipulated to produce different types of MRIs. These inclue:

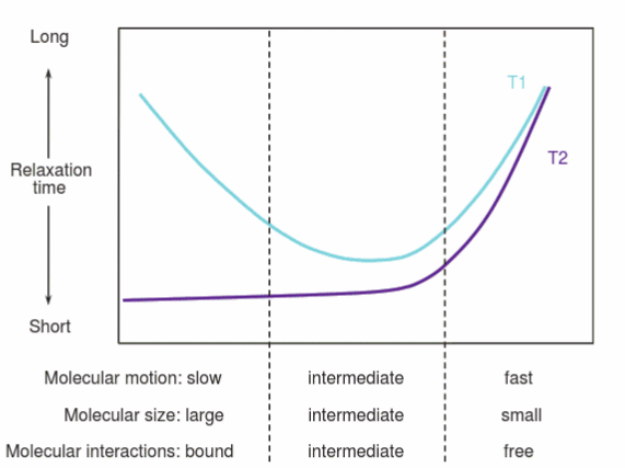


Figure 2.1: T1 and T2 relaxation comparison and the factors affecting differing relaxation times [7].

- Time of Repetition (TR): The time between excitation RF pulses, B_1, applied to B_0.
- Time of Echo (TE): The time between the excitation RF pulse and the returning signal echo's peak.
- Time of Inversion (TI): The time between a 180-degree inversion RF pulse and the 90-degree excitation RF pulse.

Spin echo (SE), gradient echo (GE), and inversion recovery (IR) are the main pulse sequences that use different combinations of TR, TE, and TI to produce images with different contrasts. For example, the TE and TR can be adjusted to create T1-weighted images and T2-weighted images.

2.0.2.1 Gradient echo

GE is a sequence that uses a magnetic field gradient to create an signal echo. After a 90-degree FID inducing pulse, the gradated magnetic field rapidly dephases the FID until an equal and opposite gradated magnetic field is applied to rephase the FID. The application of the two gradients induces the GE with relaxation of FID dictated by tissue and static field inhomogeneities [32], which is detected, recorded, and used in image reconstruction. T1-weighted spoiled gradient echo sequence is used in this study for the structural images defining GM, WM, and brain structures in which T1-weighted sequenced can be delivered and images acquired rapidly [32], detailed in Section 4.3.

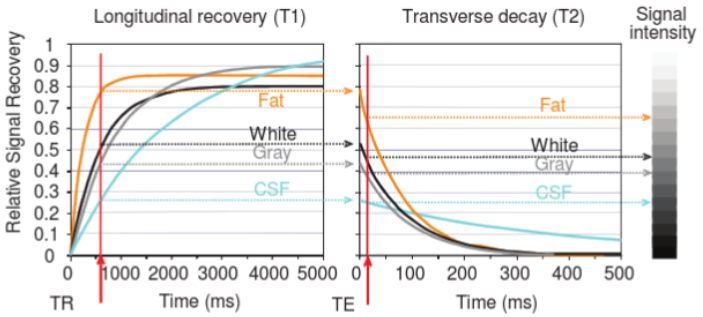
2.0.2.2 Spin echo

SE is a sequence in which a sample is initially excited and an FID is induced with a 90-degree pulse, followed by a 180-degree pulse that produces an echo for measurement. The initial excitation creates the transverse magnetisation that immediately begins to decay, and the following refocusing pulse inverts the system and induces phase coherence and peak echo amplitude at TE. The system then decays through usual mechanisms towards the restoration of equilibrium. Multiple 180-degree pulses can be applied subsequently to obtain a series of echoes that diminish according to the T2 relaxation time of the sample.

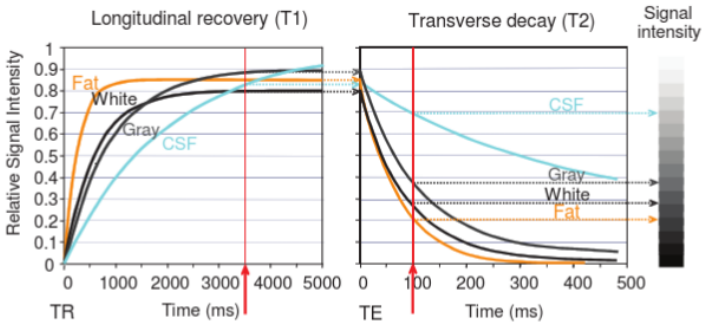
T1-weighting: One particular SE sequence used extensively in medical MRI is the T1-weighted sequence which exploits T1 characteristic differences in the sample to achieve significant contrast in the resultant images. This sequences uses a short TR in which different tissue types have vastly different signals, and pair it with a short TE to capture this difference. This combination of TR and TE is shown in Figure 2.2a, where the tissues with different molecular structures respond to the excitation RF pulse along different growth curves. T1-weighted images create high-resolution structural images with high levels of contrast between GM, WM, and CSF.

T2-weighting: T2-weighting exploits T2 characteristic differences in tissues by combining long TR and TE times. The long TR time allows the signals to recover sufficiently, and a long TE allows the signals to decay according to spin-spin relaxation. The resultant growth and relaxation curves are shown in Figure 2.2b. T2-weighted images have high soft tissue contrast, although the long TE increases image noise.

It is shown graphically in Figures 2.2a and 2.2b how different choices of TR and TE can have a significant influence on the image contrast of a particular scan. Therefore, careful selection of these times and their combinations are needed to acquire images that are clinically useful and informative.



(a) T1-weighted longitudinal recovery (L) and transverse decay (R) for 4 brain tissues.



(b) T2-weighted longitudinal recovery (L) and transverse decay (R) for 4 brain tissues.

Figure 2.2: Signal recovery as dictated by a combination of TI and TE [7]

2.0.2.3 Inversion recovery

Inversion recovery (IR) is a sequence used to suppress the response of specific tissues according to T1 characteristics. The sequence involves an inversion pulse of 180-degrees to invert the net magnetisation after which signal recovery begins as the magnetisation returns to equilibrium. A 90-degree pulse is applied after a given inversion time (TI) to transfer the recovered signal back to the transverse plane generating the FID. A secondary 180-degree pulse is applied at $TE/2$ to retrieve the signal echo for reading at TE. Dependant on the TI selected, the signal from tissues with particular molecular structures can be nulled, changing the contrast to highlight various regions of interest.

Fluid-attenuation inversion recovery (FLAIR): FLAIR imaging nulls signal of fluids with long T1 relaxation times. The TI for a FLAIR sequence is selected so when the 90-degree pulse is applied, there is no M_z signal to send to the transverse plane, thus effectively nulling the signal

[44]. The FLAIR sequence is shown in Figure 2.3. While grey and white matter are difficult to differentiate in FLAIR images [32], the images are very sensitive to small vessel disease, which manifest as white matter hyperintensities. Therefore, this thesis utilised T2 FLAIR images to identify white matter hyperintensities.

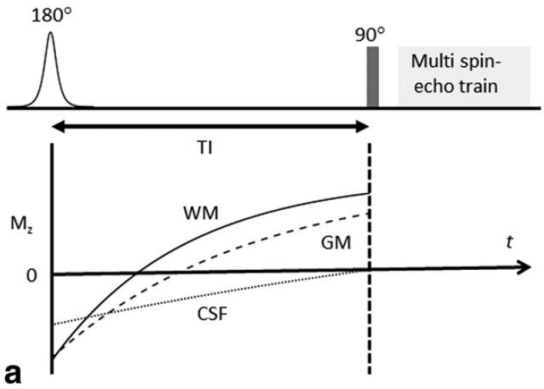


Figure 2.3: The sequence of RF pulses that form a FLAIR sequence and the relative recovery of 3 brain tissues. 90-degree RF pulse applied when CSF signal is 0 to acquire signal from only WM and GM [44].

Chapter 3

Algorithm training

Algorithm training motivation

WMHs are potentially valuable biomarkers for a number of diseases, including PD. To date, the accepted gold-standard method of identifying WMHs in MRI scans is manual detection by an expert grader [24]. This time-intensive process is not completely free from possible errors, however, and is practically unusable in large cohorts with many scans that require WMH detection. Therefore, an automated method of WMH detection is desirable, especially one that is robust and accurate. There are currently a number of automated WMH detection methods available for use, however, these methods are accompanied by particular advantages and disadvantages dependant on a myriad of factors. For instance, methods specifically designed to detect MS lesions may not be optimised for WMH detection in other diseases due to the differences in sharpness of the edge of the hyperintensities [19]. While the selection of an identification algorithm based only on published literature is an option for identifying WMHs in PD [19, 31, 56], in this thesis I directly applied and tested four published algorithms in order to determine the optimal method for the current Parkinson's cohort under investigation.

3.1 Training parameters

3.1.1 Training cohort

Here, I have set out to optimise and test four WMH algorithms. To do so, I manually traced WMHs on a group of PD participants under the supervision of a neuroradiologist. This set of manually

defined WMHs became the 'gold-standard' used to test the outputs of the four algorithms. A subset of the full study cohort was used as the training cohort since restricting the number of subjects in the training stage of this study reduced the time required to train and test the algorithms while maintaining adequate accuracy. Initially, the training cohort comprised 20 subjects with subjectively high WMH burden and were selected by global WMH volume as calculated by an initial run of a WMH identification algorithm. High burden subjects were used in the training cohort because it is suggested that high WMH burden is easier to visually identify, and an increased number of WMH voxels provides the algorithms with more learning opportunities in training to apply to new subjects, resulting in more accurate WMH detection when applied to non-training subjects [19, 23].

3.1.2 Creation of manual masks

To assess the performance of the algorithms under investigation, WMH masks were required as a 'gold-standard' benchmark of performance. I manually identified WMHs on FLAIR images of the training subjects using FSLview (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslView). Dr. Ross Keenan, a neuroradiologist, supervised and ensured accurate 'gold-standard' masks were created. Care was taken when creating the manual masks to ensure false positives were not included, and WMH clusters smaller than 5 mm in diameter were excluded per the European Task Force on Age-related White Matter changes recommendations [57]. These very small lesions are said to be caused by image noise and partial voluming effects and thus should be excluded from the WMH masks [57]. I took particular care to avoid scoring the septum pellucidum as the signal can mimic true WMHs. At the first review, the manual masks I created were advised to be too liberal and included false positives, particularly adjacent to the ventricles. These masks were revised and approved by Dr. Keenan. For the remainder of this thesis, I will refer to these manually traced WMH masks as 'manual masks'.

3.1.3 Performance measures for algorithm selection

3.1.3.1 Performance measure terminology

To understand the performance measures used to assess algorithm ability to identify WMHs it is essential to understand the terminology and definitions of the variables the metrics employ. The variables used in this study are as follows:

- **True positive (TP):** a voxel assigned to the WMH mask by an algorithm that is also included in the manual mask.
- **True negative (TN):** a voxel excluded from the WMH mask by an algorithm that is also excluded from the manual mask.
- **False positive (FP):** a voxel assigned to the WMH mask by an algorithm but is excluded from the manual mask.
- **False negative (FN):** a voxel excluded from the WMH mask by an algorithm but is included in the manual mask.
- **Detection error (DE):** the sum of voxels in false positive and negative clusters detected by the algorithm.
- **Mean total area (MTA):** the average total WMH volume by the manual mask and algorithm output.

3.1.3.2 Performance measurement metric

The Dice Similarity Index (SI) was used to quantify the similarity between the manual masks and the algorithm produced masks. SI compares the two masks on a voxel-wise level (equation 3.1). That is, SI is a ratio of TP, to TP, FP, and FN voxels between the manual and algorithm-generated masks. Secondary metrics included false negative rate (FNR), false positive rate (FPR), and detection error rate (DER) (equations 3.2, 3.3, 3.4, respectively). These three similarity metrics were chosen since a "...good balance through all DSC[Dice Similarity Coefficient], sensitivity, specificity and precision scores" [39] is needed for a learning algorithm to be considered successful. The cal-

culation of performance measures was carried out in FLAIR space for all algorithms to ensure consistency and legitimacy in comparison.

$$SI = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.1}$$

$$FNR = 1 - \frac{TN}{TN + FP} = 1 - specificity \tag{3.2}$$

$$FPR = 1 - \frac{TP}{TP + FN} = 1 - sensitivity \tag{3.3}$$

$$DER = \frac{DE}{MTA} \tag{3.4}$$

The four algorithms investigated in this study produce WMH probability maps; that is, each voxel value represented the likelihood of a region being a WMH from 0-1. In order to compare WMH probability maps to the manual masks, one needs to define a threshold to produce a binary mask. The threshold value is user defined and can vary dependant on the algorithm under investigation. For the remainder of this thesis, I will refer to binary, thresholded WMH probability maps as 'algorithm output masks'.

The following sections present and assess the performance of the 4 individual identification algorithms chosen for investigation. For each algorithm, I present an overview of how it works, the required training data and preprocessing steps required, the results of the algorithm, and any adjustments made in an attempt to improve the results. The algorithms I assessed were BIANCA, LPA, LGA, and UBO Detector [19, 23, 46, 47].

3.2 Brain Intensity AbNormality Classification Algorithm: BIANCA

3.2.1 BIANCA algorithm overview

BIANCA is an algorithm developed in 2016 [19] and made available through FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BIANCA). It is a supervised, automatic method for WMH detection based on k-nearest neighbour (k-NN) classification. According to a k-NN algorithm, the classification of a voxel of interest x is dictated by the k neighbours and is probabilistically dependant on the relative classifications of the neighbours. $k = 40$ is utilised by BIANCA based on initial performance assessment and literature [19, 51]. Training data in the form of manually identified WMHs is required for k-NN classification as the algorithm uses the positive instances in the manual masks as knowledge for the classification outputs of the algorithm. WMH identification performed by BIANCA depends heavily on the spatial location of the data point of interest as well as intensity. Since WMHs are only expected to be located in the WM, a higher probability can be given to x that falls with WM. BIANCA can be run such that a linear scaling factor is used to restrict the algorithm to using study-specific training data, improving the accuracy of WMH identification. After BIANCA is run, a post-processing step in which the probabilistic output maps of BIANCA are made binary by excluding any classified data under a user-controlled threshold. The algorithm output masks can then be compared to the manual masks and similarity measures calculated for assessment and comparison to other tested algorithms.

BIANCA was trained and optimised on the 20 selected training subjects. Unbiased training and assessment of performance for any given training specifications and parameters are of great importance. Therefore, BIANCA, uses a leave-one-out (LOO) cross-validation component in which one training subject is removed from the training data set, cohort size = n, and BIANCA is then trained using the remaining $n-1$ subjects before the resultant algorithm is applied to the one test subject who was excluded from training. The WMH probability map for each subject in the testing data set is, therefore, independent of training data.

More detailed workings of BIANCA can be found in the 2016 paper by Griffanti et al. [19].

3.2.2 BIANCA data preparation

BIANCA requires two input images for segmentation I used T1 and T2 FLAIR images in this study. One of the images must be brain extracted and both must be registered to a common space. Four combinations of brain extracted T1 or T2 FLAIR images in either T1 or T2 FLAIR space were trailed initially to determine which combination produced better results (i.e. higher similarity between manual masks and algorithm-derived masks). These four combinations are as follows; FLAIR space with brain extracted FLAIR image, FLAIR space with brain extracted T1 image, T1 space with brain extracted FLAIR image, and T1 space with brain extracted T1 image. FSL FMRIB's Linear Image Registration Tool (FLIRT) was used for registration of images to different base spaces. SPM12's registration check tool was used to ensure the registration of both T1 image to FLAIR space and FLAIR image to T1 space was accurate.

FSL Brain Extraction Toolbox (BET) was used originally for brain extraction of T1 and FLAIR images, however, the brain extraction of FLAIR images were not truly representative of the brain, namely eyes and skull were included in many brain-extracted images. This poor performance could be due to the lower resolution of the FLAIR images compared to T1 images 32 z slices compared to 170 z slices, respectively. To combat this, binary brain masks of T1 images were produced by FSL BET and multiplied with the FLAIR images in FLAIR space, resulting in far more accurate brain-extracted images.

3.2.3 Initial BIANCA results

BIANCA was run on the 20 training subjects using default optimised specifications from the 2016 paper that first presented BIANCA by Griffanti et al. [19]. For this initial training, I used a post-processing threshold of 0.95. Table 3.1 shows the results from this initial training.

Table 3.1: BIANCA results comparing input images in different spaces at post-processing threshold = 0.95. FLAIR space, brain extracted FLAIR images resulted in best performance measures. Optimal performance is indicated in bold.

Trial	SI	Voxel FPR	Voxel FNR	DER
Trial 1: FLAIR space, brain extracted FLAIR	**0.69 ± 0.02**	**0.06 ± 0.01**	**0.44 ± 0.02**	**0.07 ± 0.01**
Trial 2: FLAIR space, brain extracted T1	0.45 ± 0.17	0.13 ± 0.13	0.65 ± 0.18	0.11 ± 0.11
Trial 3: T1 space, brain extracted FLAIR	0.57 ± 0.04	0.08 ± 0.02	0.56 ± 0.04	0.13 ± 0.04
Trial 4: T1 space, brain extracted T1	0.37 ± 0.04	0.11 ± 0.03	0.74 ± 0.04	0.20 ± 0.06

Results presented as mean values ± standard deviation. Legend: SI = Similarity Index; FPR = False positive rate; FNR = False negative rate; DER = Detection error rate.

All 4 initial trials, which used a conservative threshold of 0.95, had a relatively high FNR and a low FPR, (Table 3.1). Ideally, both false negative and positive rates would be as low as possible for the most accurate results; a high FNR indicates the algorithm has failed to detect the lesions identified in the manual masks, and a high FPR indicates the algorithm has incorrectly identified regions as WMHs when they haven't been manually identified as WMHs.

Trial 1 (FLAIR space, brain extracted FLAIR) had the highest SI for the post-processing threshold of 0.95, perhaps due to the minimal registration of images to new space needed in the data preparation, and thus was selected as the image registration scheme to be subsequently used. After establishing the input data scheme, I tested the effect of different post-processing thresholds. 20 equally spaced thresholds between 0-1 were tested, with the magnitude of SI, FPR, FNR, and DER compared across all thresholds (Table 3.1). The threshold value 0.8 was selected as it produced the highest mean $SI = 0.76 \pm 0.02$, although consideration of voxel FPR and FNR, and DER were considered in the selection of the optimal threshold value.

Table 3.2: BIANCA results at various post-processing thresholds, 20 training subjects. Threshold = 0.8 resulted in the best performance measures. Optimal performance is indicated in bold.

Threshold	SI	Voxel FPR	Voxel FNR	DER
0.05	0.28 ± 0.03	0.83 ± 0.02	0.02 ± 0.00	0.03 ± 0.00
0.1	0.36 ± 0.03	0.77 ± 0.02	0.03 ± 0.00	0.04 ± 0.00
0.15	0.43 ± 0.03	0.71 ± 0.03	0.04 ± 0.01	0.05 ± 0.00
0.2	0.48 ± 0.03	0.67 ± 0.03	0.06 ± 0.01	0.06 ± 0.01
0.25	0.52 ± 0.03	0.62 ± 0.03	0.07 ± 0.01	0.06 ± 0.01
0.3	0.56 ± 0.03	0.58 ± 0.04	0.09 ± 0.01	0.07 ± 0.01
0.35	0.61 ± 0.03	0.51 ± 0.04	0.11 ± 0.01	0.07 ± 0.01
0.4	0.63 ± 0.03	0.49 ± 0.04	0.12 ± 0.01	0.07 ± 0.01
0.45	0.67 ± 0.03	0.43 ± 0.04	0.14 ± 0.01	0.06 ± 0.01
0.5	0.68 ± 0.03	0.41 ± 0.04	0.15 ± 0.01	0.06 ± 0.01
0.55	0.71 ± 0.02	0.37 ± 0.03	0.16 ± 0.02	0.06 ± 0.01
0.6	0.72 ± 0.02	0.33 ± 0.03	0.18 ± 0.02	0.06 ± 0.01
0.65	0.75 ± 0.02	0.27 ± 0.03	0.21 ± 0.02	0.06 ± 0.01
0.7	0.76 ± 0.02	0.24 ± 0.03	0.23 ± 0.02	0.05 ± 0.01
0.75	0.76 ± 0.02	0.22 ± 0.03	0.24 ± 0.02	0.05 ± 0.01
0.8	**0.76 ± 0.02**	**0.18 ± 0.02**	**0.27 ± 0.02**	**0.05 ± 0.01**
0.85	0.76 ± 0.02	0.15 ± 0.02	0.30 ± 0.02	0.05 ± 0.01
0.9	0.74 ± 0.02	0.09 ± 0.01	0.37 ± 0.02	0.06 ± 0.01
0.95	0.69 ± 0.02	0.06 ± 0.01	0.44 ± 0.02	0.07 ± 0.01

Results presented as mean values ± standard deviation. Legend: SI = Similarity Index; FPR = False positive rate; FNR = False negative rate; DER = Detection error rate.

3.2.4 BIANCA algorithm improvement

After optimising the input data and threshold cut-offs, other parameters were manipulated to improve the performance of BIANCA. Algorithm performance improvement was pursued with inclusion of additional considerations for spatial weighting (sw) and patch size. Sw uses spatial data

to give different weightings to specific voxels dependant on their position and hence the likelihood of being a WMH. The default $sw = 1$, so an increase in sw will have a larger effect on voxels deemed to be in a position likely to be a WMH. Sw values of 2 and 3 were included in the improvement investigation. A patch in the BIANCA call is an option based on the intensity data of a voxel. By classifying voxels based on neighbourhoods of voxels equal to the patch size instead of individually the algorithm is less prone to voxel false positive and false negative detection [19]. Patch sizes of $3, 4$, and 9 were tested.

An exclusion mask applied to the BIANCA output mask so that the output only included hyperintensities that appear in WM, which potentially reduces the number of false positives detected. BIANCA has a built-in exclusion mask creation script that excludes cortical GM and a number of sub-cortical nuclei and structures from the algorithm output. Alternatively, an input image can be masked before the algorithm is run so that only WM is included, eliminating the possibility of identification of hyperintensities outside the WM. Both masking options were trialed in the algorithm improvement testing.

Another improvement technique included using bias-corrected input images. Bias correction refers to the elimination of low-frequency signal produced by the MRI machine during image acquisition. This makes contrast across the images more consistent which is vital for an algorithm using voxel intensity to determine classification [25].

Additionally, 10 low burden and 10 high burden subjects were added to the training set to increase the amount of data available for training the algorithms. The inclusion of low burden training subjects was tested since it would introduce learning instances for low burden subjects, thus limiting overestimation of the algorithm trained solely on high burden subjects where large volumes of WMHs were consistent across training data but not representative of the entire cohort.

3.2.5 Improved BIANCA results

$sw = 2, 3$ and $patch = 2, 3, 4, 9$ were added to the original BIANCA call individually and in conjunction, and the combination $sw = 2$ and $patch = 3$ was found to improve the original BIANCA with only a marginal increase in the mean SI value, and was therefore not included in subsequent optimisation steps.

Assessment of the improvement made by increased training subjects, inclusion masks, and bias-corrected input images were combined. I found that increasing the training set by all 20 additional subjects improved the performance more than just including the 10 low burden additional subjects. The optimised performance of BIANCA was achieved when using bias-corrected input images in FLAIR space using 40 training subjects with no sw or patch. The BIANCA output mask was calculated with a threshold = 0.8, and a WM inclusion mask was applied to the BIANCA output. The mean dice coefficient calculated under these training specifications was $SI = 0.80 \pm 0.01$ (Table 3.3).

Table 3.3: Improved BIANCA outputs, optimal post-processing threshold reported for each trial only. Optimal performance is indicated in bold.

Training specification	SI	Voxel FPR	Voxel FNR	DER	Threshold
20	0.76 ± 0.02	0.18 ± 0.02	0.27 ± 0.02	0.05 ± 0.01	0.8
30 (low burden)	0.76 ± 0.01	0.18 ± 0.02	0.27 ± 0.02	0.06 ± 0.01	0.85
40	0.77 ± 0.01	0.21 ± 0.02	0.22 ± 0.02	0.05 ± 0.01	0.8
40 (BC)	0.80 ± 0.01	0.19 ± 0.02	0.19 ± 0.01	0.03 ± 0.00	0.85
40 (BC, masked output)	**0.80 ± 0.01**	**0.16 ± 0.02**	**0.21 ± 0.01**	**0.03 ± 0.00**	**0.8**
40 (BC, masked FLAIR input)	0.80 ± 0.01	0.17 ± 0.02	0.21 ± 0.01	0.03 ± 0.00	0.8

Results presented as mean values ± standard deviation. Legend: SI = Similarity Index; FPR = False positive rate; FNR = False negative rate; DER = Detection error rate; BC = Bias-corrected.

3.2.5.1 BIANCA results comparison to literature

My optimised BIANCA performance on the training data set was compared to published literature. Firstly, comparison to the results in Griffanti et al. (2016) showed that the optimised parameters for BIANCA (with bias-corrected images in FLAIR space and a post-processing WM inclusion mask)

outperformed the metrics of BIANCA on both published data sets with SIs of 0.76 and 0.52 [19]. A paper by Ling et al. (2018) optimised BIANCA for two datasets (n=90, 66) and the performance measures achieved, $SI = 0.79$ and $SI = 0.80$ respectively, are comparable to our performance measures [30]. Thus, the optimised BIANCA algorithm here was approximately equivalent to the high performance reported in other disease cohorts.

3.3 Lesion Growth Algorithm: LGA

3.3.1 LGA algorithm overview

Lesion Growth Algorithm (LGA) is an unsupervised classification algorithm in the Lesion Segmenting Toolbox (LST) toolbox for SPM. The toolbox provides an automated method of identifying WMHs originally created for use in Multiple Sclerosis (MS). LGA requires inputs of raw T1 and T2 FLAIR images, and only the initial threshold, κ, need be selected.

LGA operates in T1 space, and can essentially be broken into 3 steps. In the pre-processing step, the T1 image is segmented into grey matter (GM), WM, and cerebrospinal fluid (CSF). The FLAIR image is bias-corrected and registered to T1 space along with a standard SPM WM probability map. In the second step, lesion belief maps for the 3 tissues are created by identifying hyperintensities based on FLAIR voxel intensity and spatial location. These initial lesion belief maps are summed, $B = b_1, \ldots, b_n$, and the GM belief map, B_{GM}, is used to create a seed for the lesion mask to grow, as increasing voxel values of B_{GM} indicate that the voxel is more likely to be included in a WM lesion (see Schmidt et al., (2012) section 'Lesion belief maps and initialization' for detail). The value of an individual voxel i in the GM belief map, $b_{GM,i}$, is used to determine the class of the corresponding individual voxel in the initial lesion map, $l_{init,i}$, by equation 3.5. Hence $L_{init,1} = l_{init,1}, ..., l_{init,i}$ is the binary map used to initiate the iterative growth of the WMH map.

$$l_{init,i} = 1 \leftrightarrow b_{GM,i} > \kappa \tag{3.5}$$

The third step grows $L_{init,i}$ towards B via an iterative process. The algorithm identifies voxels adjacent to any voxel included in $L_{init,1}$ and tests it to either accept or reject the voxel under consideration from the growing lesion map. The adjacent voxel under consideration is assigned a probability value of being a lesion voxel with considerations for size and shape of gamma distribution describing the lesion map and utilises a Markov random field (MRF) to assure the probabilistic value assigned to the voxel of interest depends only on labelled voxels in its neighbourhood. The

voxel is assigned a probability value according to the equation 3.6:

$$\pi_i^{Les} = Pr(z_i = Les) = min\left(1, \frac{P_{Les}(y_i|\hat{\alpha}^{(t-1)}, \hat{\beta}^{(t-1)}) \cdot b_i \cdot exp(-\sum_{j\epsilon N_i}(1-\pi_j^{Les}))}{P_{other}(y_i|\hat{\theta}^{(t-1)}) \cdot exp(-\sum_{j\epsilon N_i}\pi_j^{Les})}\right) \tag{3.6}$$

Here, N_i is the first order neighbourhood of the voxel of interest, i, wherein there is at least one voxel, j that has a lesion probability $\pi_j^{Les} > 0, j\epsilon N_i$. y_i is the normalised FLAIR intensity of i. $Pr(z_i = Les)$ is the probability value of the voxel belonging to the lesion map. P_{Les} is the gamma distribution density function for the lesion class with parameters α for size and β for shape. P_{other} is the combined GM, WM, CSF Gaussian distribution. The iterative process will halt when the preset maximum number of iterations is reached or the largest new lesion probability is < 0.01. More thorough details of this algorithm are provided in the 2012 paper by Schmidt et al. [47].

3.3.2 Initial LGA results

LGA v 2.0.15 was used in this study. Initial values of $\kappa = 0.10, 0.25, 0.30, 0.35, 0.40$ were investigated and inputs of raw T1 and FLAIR were used. As with BIANCA, LGA output masks were compared to the manual masks based on the defined performance measures. The LGA probability maps were in T1 space, thus needed to be registered to FLAIR space to calculate the equivalent algorithm performance measures, and FSL FLIRT was used to register the WMH probability maps to FLAIR space. Thresholds between 0-1 were then used to assess algorithm performance, and a post-processing threshold = 0.05 was found to consistently be the threshold that resulted in the highest SI value across all κ values; this threshold was used to compare the different iterations of the algorithm. $\kappa = 0.1$ was the best value to use, and resulted in $SI = 0.63 \pm 0.02$ (Table 3.4).

3.3.3 LGA algorithm improvement and results

An attempt to improve the performance of LGA involved coregistering the raw T1 image to FLAIR space for the T1 input. Again, a threshold value = 0.1 resulted in the highest SI, however, the results did not outperform the original raw T1 input images (Table 3.5). The coregistration of the T1 image was successful for all training subjects, however, LGA did not work on one subject after the coregistration of input images. This subject was excluded from the subsequent calculation of performance measures of LGA using coregistared T1 images.

Table 3.4: LGA results: Input of raw FLAIR and T1 images calculated in FLAIR space. Optimal performance is indicated in bold.

κ	SI	Voxel FPR	Voxel FNR	DER
0.10	**0.63 ± 0.02**	**0.20 ± 0.02**	**0.47 ± 0.03**	**0.11 ± 0.02**
0.25	0.59 ± 0.02	0.13 ± 0.02	0.54 ± 0.03	0.11 ± 0.02
0.30	0.57 ± 0.03	0.12 ± 0.02	0.57 ± 0.03	0.11 ± 0.02
0.35	0.55 ± 0.03	0.12 ± 0.02	0.59 ± 0.03	0.12 ± 0.03
0.40	0.53 ± 0.03	0.11 ± 0.02	0.61 ± 0.03	0.12 ± 0.03

Results presented as mean values ± standard deviation. Legend: SI = Similarity Index; FPR = False positive rate; FNR = False negative rate; DER = Detection error rate; κ = parameter determining how conservative initial lesion map.

Table 3.5: LGA test results presenting input images of raw FLAIR and coregistered T1. Performance measures calculated in FLAIR space. Optimal performance is indicated in bold.

κ	SI	Voxel FPR	Voxel FNR	DER
0.05	0.60 ± 0.03	0.16 ± 0.03	0.53 ± 0.03	0.12 ± 0.03
0.10	**0.59 ± 0.03**	**0.10 ± 0.02**	**0.55 ± 0.03**	**0.10 ± 0.02**
0.25	0.53 ± 0.03	0.06 ± 0.01	0.63 ± 0.03	0.12 ± 0.03
0.30	0.51 ± 0.03	0.05 ± 0.01	0.65 ± 0.03	0.13 ± 0.03

Results presented as mean values ± standard deviation. Legend: SI = Similarity Index; FPR = False positive rate; FNR = False negative rate; DER = Detection error rate; κ = parameter determining how conservative initial lesion map.

The number of maximum iterations was also increased from 50 to 150 as some subjects map growths were stopped since they reached 50 iterations instead of reaching a point where the probability values of all voxels dropped below 0.01. These alterations of LGA improved the FPR detection, however, consideration must be given to the fact that only 19 of 20 subjects could be used to calculate the performance measures. With this considered, the original result of LGA performance using $\kappa = 0.1$ and raw T1 and FLAIR T2 input images was used as the optimal performance of the algorithm, and used for comparison to the other tested algorithms.

Table 3.6: LGA similarity index comparison to literature values. Optimal performance is indicated in bold.

Paper	κ	Training Subjects	SI
Our results	0.1	20 PD	0.63
Schmidt et al. (2012)	**0.3**	**53 MS**	**0.75**
Griffanti et al. (2016)	0.2	85 AD	0.69
Rachmadi et al. (2017)	0.13	20 AD	0.30

Legend: SI = Similarity index; PD = Parkinson's disease; AD = Alzheimer's disease; MS = Multiple sclerosis; κ = parameter determining how conservative initial lesion map.

3.3.3.1 LGA result comparison to literature

The performance of LGA in our training data set was inconsistent with the literature published for other studies which reported a range of SIs. In the 2012 paper by Schmidt et al. where this method was first presented, the average SI for a cohort of 53 MS patients was $SI = 0.75$ when using a value of $\kappa = 0.3$ [47]. The 2016 paper by Griffanti et al. reported $SI = 0.69$ at $\kappa = 0.2$ when used on a cohort of 85 individuals with AD [19]. A more recent paper by Rachmadi et al., 2017 reported an average SI of 0.29 at $\kappa = 0.13$ on an AD cohort of 20 [39]. These results have been tabulated in Table 3.6. This variation in SI across studies may be due to the inherent differences between the sampled studies, such as the cohorts used in terms of the number of patients and cohort disease, and the inconsistencies in the MR images chiefly caused by different acquisition machines and MRI protocols. All of these factors could have impacted the performance of LGA and the resultant SI values calculated.

3.4 Lesion Probability Algorithm: LPA

3.4.1 LPA algorithm overview

Lesion Probability Algorithm (LPA) is a supervised classification lesion segmenting algorithm in the LST toolbox v 2.0.15 of SPM. LPA uses LGA and its default settings, as detailed above in 3.3.1, to estimate the location of the WM lesions and then applies this learned information to different study cohorts. LPA uses prior WMH knowledge as determined by LGA trained using a cohort of 53 MS patients with high WMH burden, a total lesion volume (TVL) > 10 ml, and lesion probability maps for each training individual are calculated. The only inputs required for LPA are T2-weighted FLAIR images since the inclusion of T1-weighted image is only needed for the creation of reference lesion maps in LGA, which has already been completed in the (independent) MS training cohort.

LPA first runs a bias correction step on the FLAIR images to regulate the intensity features of the images. Then lesion probability maps are created iteratively in the same way as LGA except the reference lesion maps are calculate by LGA in the training cohort instead of per subject within the cohort of interest. LPA produces the linear predictor, $\hat{\eta}_i$, by combining posterior means as in equation 3.7, wherein β_0 is the intercept, x_i is the lesion belief map for the ith voxel, β_1 is the effect of the lesion belief map, and γ_i models the spatial effect.

$$\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\gamma}_i \tag{3.7}$$

Lesion segmentation is then executed in voxel i by

$$\hat{\pi}_i = \frac{exp(\hat{\eta}_i)}{1 + exp(\hat{\eta}_i)} \tag{3.8}$$

There is a post-processing step included in LPA that removes any positively identified WMH if it is $< 0.015ml$ in volume. LPA reportedly works well with FLAIR input alone as additional modality images require additional pre-processing of coregistration to the same, space which is subject to error, particularly for subjects with high WMH burden [46].

More detail on LPA can be found in the 2017 thesis by Schmidt et al. [46].

3.4.2 LPA data preparation

LPA requires only the FLAIR image input and needs no initial parameters selected. Additionally, there is an option of adding reference images (structural T1-weighted) to the input data. LPA was therefore trialled both with and without reference images. Coregistration of input images only occurs when a T1 image is included in the algorithm input. The output of LPA is a non-binary lesion probability map that required thresholding to produce a binary WMH mask. A range of threshold values between 0-1 were tested to find the best one for use in the training data set (i.e. highest SI to the study 'gold-standard' manual masks).

3.4.3 LPA results

LPA performed better when only a FLAIR image was used as the input image, perhaps due to the coregistration error highlighted by Schmidt et al. [46]. The highest SI value achieved by LPA was $SI = 0.67 \pm 0.03$ when the LPA output mask was calculated at a post-processing threshold $= 0.35$. In terms of improving LPA's performance, bias-corrected inputs were not considered for improvement of LPA because a bias correction is included in the algorithm. There is no training data input required for the algorithm, so increasing the test inputs from 20 to 40 was unnecessary for LPA.

When FLAIR was the only image input, the LPA probability maps are produced in FLAIR space. When T1 images were included as the input and coregistration occurred, the LPA probability maps are produced in T1 space. In the latter case, the probability maps were registered to FLAIR space using FSL FLIRT for performance measure calculation.

Table 3.7: LPA test results at different post-processing thresholds. Results presented for trial exclusively using FLAIR images, and FLAIR and T1-weighted images. Optimal performance is indicated in bold.

Threshold	SI	Voxel FNR	Voxel FPR	DER
0.05 FLAIR only	0.61 ± 0.04	0.45 ± 0.06	0.16 ± 0.02	0.06 ± 0.01
0.1 FLAIR only	0.62 ± 0.04	0.44 ± 0.06	0.17 ± 0.02	0.07 ± 0.02
0.15 FLAIR only	0.63 ± 0.04	0.41 ± 0.06	0.18 ± 0.02	0.07 ± 0.02
0.2 FLAIR only	0.65 ± 0.04	0.38 ± 0.05	0.20 ± 0.02	0.07 ± 0.02
0.25 FLAIR only	0.67 ± 0.03	0.35 ± 0.05	0.23 ± 0.02	0.07 ± 0.02
0.3 FLAIR only	0.67 ± 0.03	0.33 ± 0.05	0.25 ± 0.02	0.07 ± 0.02
0.35 FLAIR only	**0.67 ± 0.03**	**0.30 ± 0.05**	**0.27 ± 0.03**	**0.08 ± 0.02**
0.4 FLAIR only	0.67 ± 0.03	0.28 ± 0.05	0.30 ± 0.03	0.08 ± 0.02
0.45 FLAIR only	0.67 ± 0.03	0.26 ± 0.05	0.33 ± 0.03	0.08 ± 0.02
0.5 FLAIR only	0.66 ± 0.03	0.24 ± 0.05	0.35 ± 0.03	0.08 ± 0.02
0.55 FLAIR only	0.66 ± 0.02	0.22 ± 0.05	0.38 ± 0.03	0.08 ± 0.02
0.6 FLAIR only	0.64 ± 0.02	0.20 ± 0.05	0.42 ± 0.03	0.09 ± 0.01
0.65 FLAIR only	0.62 ± 0.02	0.19 ± 0.05	0.46 ± 0.03	0.09 ± 0.01
0.7 FLAIR only	0.59 ± 0.03	0.15 ± 0.04	0.51 ± 0.03	0.10 ± 0.02
0.75 FLAIR only	0.55 ± 0.03	0.14 ± 0.04	0.56 ± 0.03	0.12 ± 0.03
0.8 FLAIR only	0.50 ± 0.03	0.12 ± 0.04	0.63 ± 0.03	0.13 ± 0.03
0.85 FLAIR only	0.37 ± 0.03	0.08 ± 0.03	0.76 ± 0.02	0.14 ± 0.03
0.9 FLAIR only	0.30 ± 0.02	0.06 ± 0.02	0.82 ± 0.02	0.15 ± 0.04
0.95 FLAIR only	0.23 ± 0.02	0.04 ± 0.02	0.87 ± 0.02	0.16 ± 0.04
0.05 FLAIR and T1	0.34 ± 0.04	0.70 ± 0.03	0.57 ± 0.04	0.11 ± 0.03
0.1 FLAIR and T1	0.34 ± 0.04	0.68 ± 0.04	0.60 ± 0.04	0.12 ± 0.04
0.15 FLAIR and T1	0.34 ± 0.04	0.67 ± 0.04	0.62 ± 0.04	0.13 ± 0.04
0.2 FLAIR and T1	0.34 ± 0.04	0.66 ± 0.04	0.64 ± 0.04	0.13 ± 0.04
0.25 FLAIR and T1	0.34 ± 0.04	0.65 ± 0.04	0.65 ± 0.04	0.14 ± 0.04
0.3 FLAIR and T1	0.33 ± 0.04	0.64 ± 0.04	0.67 ± 0.04	0.14 ± 0.04
0.35 FLAIR and T1	0.33 ± 0.04	0.63 ± 0.04	0.68 ± 0.04	0.14 ± 0.04
0.4 FLAIR and T1	0.33 ± 0.04	0.62 ± 0.04	0.70 ± 0.04	0.14 ± 0.04
0.45 FLAIR and T1	0.32 ± 0.04	0.61 ± 0.05	0.71 ± 0.03	0.14 ± 0.04
0.5 FLAIR and T1	0.32 ± 0.04	0.60 ± 0.05	0.72 ± 0.03	0.14 ± 0.04
0.55 FLAIR and T1	0.31 ± 0.04	0.59 ± 0.05	0.74 ± 0.03	0.14 ± 0.03
0.6 FLAIR and T1	0.30 ± 0.04	0.59 ± 0.05	0.75 ± 0.03	0.14 ± 0.03
0.65 FLAIR and T1	0.29 ± 0.04	0.58 ± 0.05	0.77 ± 0.03	0.14 ± 0.03
0.7 FLAIR and T1	0.28 ± 0.03	0.57 ± 0.05	0.78 ± 0.03	0.15 ± 0.04
0.75 FLAIR and T1	0.27 ± 0.03	0.56 ± 0.05	0.80 ± 0.03	0.15 ± 0.04
0.8 FLAIR and T1	0.25 ± 0.03	0.55 ± 0.05	0.82 ± 0.02	0.16 ± 0.04
0.85 FLAIR and T1	0.23 ± 0.03	0.55 ± 0.05	0.84 ± 0.02	0.17 ± 0.04
0.9 FLAIR and T1	0.21 ± 0.03	0.54 ± 0.05	0.86 ± 0.02	0.17 ± 0.04
0.95 FLAIR and T1	0.17 ± 0.03	0.53 ± 0.06	0.90 ± 0.02	0.20 ± 0.06

Results presented as mean values ± standard deviation. Legend: SI = Similarity Index; FPR = False positive rate; FNR = False negative rate; DER = Detection error rate.

3.5 Unidentified Bright Object Detector: UBO Detector

3.5.1 UBO Detector algorithm overview

UBO Detector is the most recent detection algorithm that was tested, being developed and released in 2017 by Jiang and colleagues [23]. UBO is a fully automated, supervised detection method that uses a combination of SPM tools for segmentation and registration and a k-NN algorithm to identify WMHs. There is a built-in post-processing step in which the global WMHs are further separated into regional WMHs; namely periventricular WMHs (PVWMHs), deep WMH (DWMH), and lobar and arterial regional WMHs. UBO Detector is unique to the other algorithms tested as it operates standardised (normalised or MNI) space.

3.5.1.1 Pre-processing

UBO takes inputs of raw FLAIR images and corregisters them to the higher resolution T1-weighted images. Tissue class probability maps (GM, WM, and CSF) are generated by segmentation of the T1-weighted images. DARTEL (Diffeomorphic Anatomical Registration Through Exponentiated Lie) is then run on the original T1-weighted image creating a subject-specific warp flow field that can be used then to register the tissue class probability maps and FLAIR images to normalised space also. This is followed by bias correction of the normalised FLAIR image. Non-brain tissues are then removed in normalised space before FSL FAST is run on the normalised FLAIR image to segment GM, WM, and CSF to effectively identify clusters of interest for WMH identification.

3.5.1.2 WMH extraction

k-NN, as described in Section 3.2.1, is then applied to the images in MNI (normalised) space, with $k = 5$, the default k-value. The training data used for UBO Detector consisted of 10 individuals in the Sydney Memory and Ageing Study (Sydney MAS), and true positive WMHs were chosen by visually assessing FAST extraction on the training set by Jiang et al. [24]. UBO Detector also offers the option of including study-specific training data to create a classification model that potentially better suited to the data set in use. The outputs of the k-NN algorithm are improved

by the inclusion of intensity and spatial location values, as detailed in the 2018 paper by Jiang et al. [24]. WMH probability maps are then created from the k-NN algorithm output using these intensity and spatial features.

3.5.1.3 Post-processing

The built-in post-processing separates the global WMH map output into sub-regions by two methods. Firstly, PVWMHs, DWMHs, and lobes are separated by a distance map method in which WMH detected within 12mm around the ventricles are identified as PVWMHs. Any WMHs that lie outside this 12mm mask, ie non-PVWMHs, are then further separated into lobes such that all WMHs identified by UBO Detector can be specifically categorised. UBO Detector is an appealing choice of WMH detector due to this additional post-processing step of segmenting the output masks due to our future interest in the WMH burden of distinct brain areas and lobes.

Further detail of UBO Detector can be found in the 2018 paper by Jiang et al. [24].

3.5.2 UBO Detector initial results

As mentioned above, UBO Detector can be run with the default parameters, or these can also be changed to better suit the cohort being segmented by creating a custom classifier. Optimised parameters for the built-in k-NN classifier were determined by a LOO cross-validation method on the built-in MS training data set [24]. Initially, UBO Detector was run with the recommended settings of k = 5 and a post-processing threshold = 0.7. Assessment of similarity measures were calculated using the binary FLAIR space mask output for consistency with the other algorithms opposed to in normalised space. Using the default algorithm settings, $SI = 0.61 \pm 0.02$ was achieved, see Table 3.8 for all performance measures.

3.5.3 UBO Detector algorithm improvement

There were a series of improvements that were tested on UBO Detector that followed from the improvements made with BIANCA. Using the custom classifier, increasing the training data to 40

subjects, and using bias-corrected input images were all pursued for improvement in performance for the same reasons as in Section 3.2.4.

3.5.4 UBO Detector improved results

The results of all the trials of UBO detector are presented in Table 3.8 with the highest SI value produced by using bias-corrected input FLAIR and T1-weighted images for 40 training subject using a custom classifier (Table 3.8).

Table 3.8: UBO test results. Optimal performance is indicated in bold.

Trial	SI	Voxel FNR	Voxel FPR	DER
20 Training, built-in classifier	0.61 ± 0.02	0.12 ± 0.02	0.52 ± 0.02	0.08 ± 0.02
20 Training, custom classifier	0.53 ± 0.04	0.08 ± 0.01	0.60 ± 0.04	0.23 ± 0.07
40 Training, custom classifier	0.64 ± 0.02	0.36 ± 0.03	0.34 ± 0.02	0.08 ± 0.01
40 Training, custom classifier, bias-corrected	**0.66 ± 0.02**	**0.34 ± 0.02**	**0.31 ± 0.02**	**0.08 ± 0.01**

Results presented as mean values ± standard deviation. Legend: SI = Similarity Index; FPR = False positive rate; FNR = False negative rate; DER = Detection error rate.

3.6 Algorithm selection and justification

In the training data, BIANCA was the WMH detection method that exhibited the strongest performance, with individual optimised algorithm performance for the four algorithms tested as follows; BIANCA SI = 0.80 ± 0.01; LGA SI = 0.63 ± 0.02; LPA SI = 0.67 ± 0.03; UBO Detector SI = 0.66 ± 0.02. The performance of BIANCA resulted in an SI comparable to those quoted in the reviewed literature. Similarly, BIANCA consistently proved the highest SIs across the 4 methods trialled, even before improvement steps were taken.

It is important to note that due to their nature, LGA and LPA posed fewer options for improvement as the inputs and initial parameters are limited. BIANCA and UBO Detector offer more opportunity to train the algorithm to the specific cohort being used. BIANCA also resulted in the best performance measures across all 4 algorithms even before any study-specific improvements had been made, and BIANCA resulted in the highest SI on an individual subject level across all training subjects and all algorithms. For these reasons, BIANCA was selected and applied to the entire cohort, using optimised algorithm parameters.

Increasing the size of the training data set was one of the major improvement technique, specifically by including low WMH burden training subjects. The training data set was increased to explicitly include low WMH burden subjects to remedy overestimation of the algorithm when applied to the entire set of MRI scans being used in this study. Overestimation manifest in the algorithm output masks as a large number of small false positive 'speckles' throughout the volume of the brain, especially in and close to cortical regions. It became clear that the inclusion of low WMH burden training subjects did not improve the performance of the algorithms as the overestimation persisted after the training cohort was increased. The overestimation was therefore removed by applying a threshold for very small lesions that were too small to be considered true positives, details of this follow in Chapter 4.

Bias-corrected input images were much more successful in improving the resultant algorithm output masks compared to the increase in the number of training subjects. The bias-corrected images

improved contrast levels between different tissue types, as well as between healthy and pathological WM. Improved contrast levels resulted in more successful k-NN algorithm WMH identification due to the nature of the algorithm and its use of voxel intensity. Testing the improvement of algorithm performance using increased training data set size and using bias-corrected images demonstrated that bias correction was a superior improvement technique, as demonstrated in the results of BIANCA (Table 3.3).

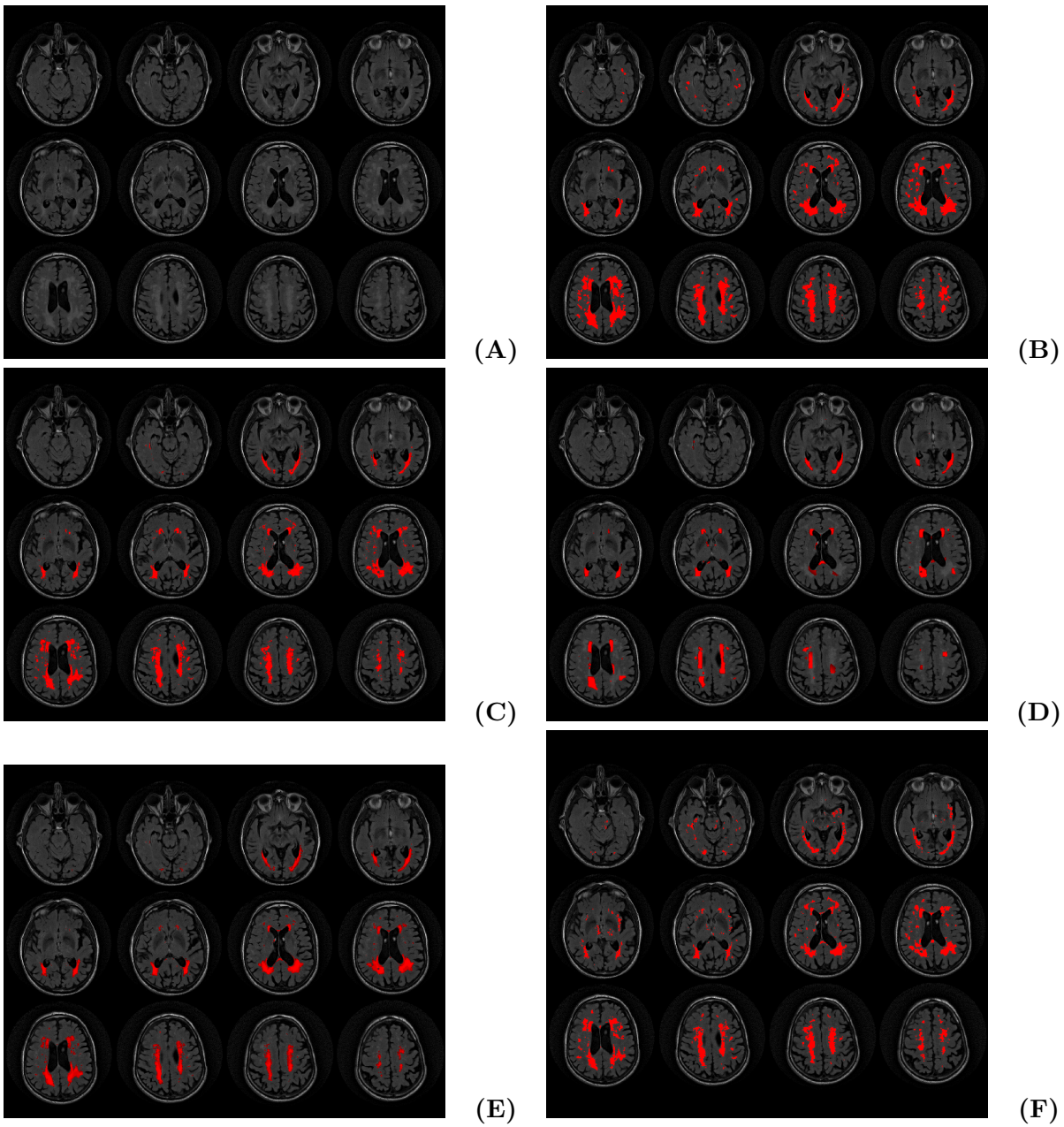


Figure 3.1: Representative FLAIR image of training subjects and results from 4 optimised algorithms. (a) Bias-corrected FLAIR image. (b)Manual mask. (c) BIANCA: Bias-corrected input images, FLAIR space, 40 training subjects, threshold = 0.80, $SI = 0.82$. (d) LGA: Raw T1 and FLAIR, $\kappa = 0.1$, threshold = 0.05, $SI = 0.55$. (e) LPA: FLAIR only, threshold = 0.35, $SI = 0.76$. (f) UBO Detector: 40 training subjects, custom classifier, threshold = 0.70, $SI = 0.71$

.

Chapter 4

Study specifications and analysis methods

This chapter presents the full longitudinal PD cohort and the analytical techniques used to investigate the relationship in question between WMHs and cognitive impairment in PD. The analysis that follows from this point uses the WMH masks produced by BIANCA using the optimal parameters and inputs determined in Chapter 3 applied to the full cohort.

4.1 Cohort

The data utilised in this study was a subset of the data in an ongoing longitudinal study that began in May 2007. Study participants were recruited from the Movement Disorders Clinic at the New Zealand Brain Research Institute, Christchurch, New Zealand. The longitudinal cohort utilised in this thesis comprised 232 participants meeting the UK Parkinson's Disease Society's criteria for idiopathic PD, with participants representative of the broad spectrum of cognitive impairment in PD. The control group was made up of 64 healthy volunteers who were matched to the average characteristics of PD participants, including age, sex ratio, and years of education.

Participants were assessed at baseline and had follow-up assessments approximately every two years after baseline, completing a rigorous neuropsychological battery, clinical assessment, and 3 T MRI scanning at each assessment. Exclusion criteria for this cohort included atypical Parkinsonian

Table 4.1: Study subjects at baseline

Cognitive group	Control	PD-N	PD-MCI	PDD	ANOVA	Overall p-value	Pairwise comparison (Tukey HSD) $>$ / $<\sim<$ 0.05
Baseline n	49	81	71	27			
Sex baseline (F:M)	16:33	29:52	17:54	4:23			
Mean age baseline	69.3 (8.4)	65.1 (8.0)	71.4 (6.5)	74.7 (6.3) e	$F(3, 224) = 15.1$	< 0.001	HC > PD-N < PD-MCI ∼ PDD
Education years baseline	13.4 (2.8)	13.0 (2.7)	12.6 (3.0)	12.4 (2.4)	$F(3, 224) = 1.2$	$= 0.329$	$HC \sim PD - N \sim PD - MCI \sim PDD$
Cognitive z score baseline	0.59 (0.39)	0.21 (0.44)	-0.86 (0.47)	-1.83 (0.52)	$F(3, 224) = 241.1$	< 0.001	$HC > PD - N > PD - MCI > PDD$
MoCA score baseline	26.9 (2.1)	26.5 (2.2)	22.7 (3.1)	16.8 (3.6)	$F(3, 221) = 111.7$	< 0.001	$HC > PD - N > PD - MCI > PDD$
Executive function domain baseline	0.62 (0.52)	0.32 (0.30)	-1.00 (0.72)	-2.02 (0.50)	$F(3, 224) = 165.0$	< 0.001	$HC > PD - N > PD - MCI > PDD$
Attention domain baseline	0.32 (0.49)	0.05 (0.47)	-0.97 (0.53)	-1.94 (0.61)	$F(3, 224) = 162.6$	< 0.001	$HC > PD - N > PD - MCI > PDD$
Memory domain baseline	0.89 (0.77)	0.26 (0.80)	-0.89 (0.71)	-1.74 (0.72)	$F(3, 224) = 101.9$	< 0.001	$HC > PD - N > PD - MCI > PDD$
Visuospatial domain baseline	0.53 (0.55)	0.22 (0.50)	-0.58 (0.72)	-1.63 (0.75)	$F(3, 224) = 91.5$	< 0.001	$HC > PD - N > PD - MCI > PDD$

Baseline study subject data, with details for each cognitive domain. Data presented as mean(SD) or ratios.
HC = Healthy control; PD-N = Parkinson's disease normal cognition; PD-MCI = Parkinson's disease mild cognitive impairment; PDD = Parkinson's disease dementia; MoCA = Montreal Cognitive Assessment.

disorder, prior learning disability, history of other neurological conditions (moderate-severe head injury, stroke, vascular dementia, major psychiatric or medical illness in the previous 6 months). Baseline MRI screening excluded subjects presenting significant non-PD pathology, including one control and three PD for severe WM disease, one PD for marked cerebral atrophy, one control for arachnoid cyst and chronic inflammatory demyelinating polyneuropathy, one control for cerebellar infarcts, and one PD for cortical infarcts. Three PD were excluded due to motion artefacts across all MR images, and three controls were excluded for meeting criteria for MCI.

Over the duration of the study, three PD subjects had cortical stroke, one PD subject was diagnosed as progressive supranuclear palsy (PSP) and one PD subject was diagnosed as multiple system atrophy (MSA). All data from those participants re-classified as having an atypical parkinsonian disorder were excluded, while images taken prior to stroke were included. The final cohort analysed in this thesis, therefore, included 207 PD and 51 Control participants that completed at least one MRI and associated cognitive testing session. A total of 585 MRI scans collected over the duration of the longitudinal study were utilised in this study, with a total of 217 PD-N, 168 PD-MCI, 42 PDD, and 158 control scans analysed, Table 4.1 provides details of these subjects. The study was approved by the regional Ethics Committee of the New Zealand Ministry of Health. All participants were given written, informed consent, with additional consent from a significant other when appropriate.

4.2 Cognitive diagnostic criteria and assessment

Comprehensive cognitive testing consistent with Movement Disorder Society Task Force level II criteria for the diagnosis of MCI was carried out, per Wood et al. (2016) [59]. PDD diagnosis was also defined by the Movement Disorders (MDS) Task Force criteria [14]. The cognitive test battery included at least two tests within 5 cognitive domains; executive function, attention and working memory, learning and memory, visuospatial performance and language. Table 4.2 presents these tests [13, 59]. PD-MCI was diagnosed when the patient's functional activities in daily life were unimpaired, but their cognitive scores were 1.5 SDs or more below normative data in at least two tests of at least one of the assessed cognitive domains, excluding language (language was excluded due to low variance). Global cognitive performance was quantified by an aggregate z score derived from the mean of the average standardised scores in the four cognitive domains listed above, excluding language. Additional cognitive tests included the Montreal Cognitive Assessment (MoCA), the Dementia Rating Scale (MDRS-2), and the Alzheimer's Disease Assessment Scale Cognition (ADAS-Cog) tests. Functional and psychiatric status was assessed with Instrumental Activities of Daily Living, Clinical Dementia Rating (CDR) status, Global Deterioration Scale, Parkinson's Disease Questionnaire (PDQ-39), Neuropsychiatric Inventory (NPI), and the Geriatric Depression Scale [2].

For brevity, for the remainder of this thesis, I will refer to the learning and memory domain as the memory domain, and the attention, working memory, and processing speed domain as the attention domain.

4.3 MRI acquisition

All images were acquired using a 3 T General Electric HDxt scanner with an eight-channel head coil. Structural MR images were obtained according to the following specifications;

T1-weighted 3D spoiled gradient recall echo (SPGE), TE/TR = 2.8/6.6 ms, TI=400 ms, flip angle=15°, acquisition matrix=256×256×170, FOV=250 mm, slice thickness=1 mm, voxel size=0.98×0.98×1.0 mm^3.

T2-weighted FLAIR PROPELLOR (motion insensitive radial k-space) 2D, TE/TR = 105/9000 ms, TI =2250 ms, echo train length=36, acquisition matrix=320×320×35, FOV=220 mm, slice thickness=3mm, gap=1.5 mm, reconstruction voxel size=0.43×0.43×4.5 mm^3.

4.4 Application of BIANCA

As discussed in Chapter 3, of the tested algorithms BIANCA produced the best performance metrics. Thus BIANCA with optimised parameters was used to create probabilistic WMH lesion masks in all study participants at each time point.

4.5 BIANCA output mask normalisation

Normalisation of the WMH output probability maps from BIANCA WMH masks was carried out by warping images in subject space to a standardised template space (in this thesis, I used the DARTEL MNI template available with CAT12). The normalisation process attempts to remove differences in brain size and shape of an individual, while at the same time maintaining subject-specific features. Normalisation of all BIANCA WMH maps was based on the deformation parameters derived from normalising the registered T1-weighted images; normalisation was carried out using CAT12, a toolbox of SPM12 (Section 4.6.1).

The data in this study were analysed both cross-sectionally and longitudinally. Therefore, I performed two different processing streams, presented below.

4.5.1 FLAIR space to T1 space

The first step in normalisation of BIANCA WMH masks for longitudinal and cross-sectional data is the removal of small, false positive clusters from the binary FLAIR output mask. Clusters of 10 voxels or smaller were eliminated from the masks as clusters under this size most likely represent false positive WMHs [57]. A cluster limit of 10 voxels was selected following the work of Wahlund et al., in which 5 mm diameter WMHs were stated to be false positive. 10 voxels in the FLAIR images is a conservative number, potentially failing to remove some false positive clusters, however, it

Table 4.2: Tests specific to cognitive domain.

Attention, Working Memory, and Processing Speed	Executive Function	Visuospatial/Visuoceptual	Learning and Memory	Language
Digits forward/backward	Action (verb) fluency	Judgement of line orientation	CVLT-II SF acquisition	Boston naming test
Digit ordering	Letter fluency (D-KEFS)	Fragmented letters	CVLT-II SF short delay (30 seconds)	DRS-2 similarities subtests
Map test (test of every day attention)	Category fluency (D-KEFS)	Rey complex figure copy	CVLT-II SF long delay (10 minutes)	ADAS-Cog (object and finger naming, commands, comprehension, spoken language and word finding difficulties)
Stroop colour reading	Category switching	Picture completion	RCF test - short delay (3 minutes)	
Stroop word reading	Trails B		RCF test - long delay (30 minutes)	
Trails A	Stroop interference (D-KEFS)			

Legend: CVLT-II = California Verbal Learning TestSecond Edition; D-KEFS = DelisKaplan Executive Function System; ROC = ReyOsterrieth complex; ADAS-Cog = Alzheimer's Disease Assessment Scale-Cog.

was chosen to ensure no true positive clusters would be removed from the BIANCA output masks. This is a trade-off between removing false positive WMHs and ensuring no true positive WMHs are removed in the process.

After false positive clusters were removed, the BIANCA WMH masks were coregistered to the T1 space. The T2 FLAIR images were coregistared to the T1 images using SPM12's Coregister: Estimate and Reslice, which is used for intra-subject registration with rigid-body model and reslicing. Coregister: Estimate and Reslice includes a smoothing step to reduce the change of local maxima in the coregistered image, increase convergence time, and smooths the cost function as much as possible (spm12/man/manual.pdf). The coregistration tool uses the T1-weighted image as a reference, the T2-weighted FLAIR image as the source image (the image which is reoriented to match the reference image), and the BIANCA WMH mask in FLAIR space is an additional image which is coregistared along with the T2 FLAIR. All other parameters used are the standard default settings. The coregistration assured that the WMH masks were aligned to T1 space, which then allowed application of the structurally-derived deformation fields to warp WMH masks into normalised space (described in Section 4.6.2).

4.5.2 Longitudinal average T1 images

At this point, the normalisation process differs slightly for longitudinal and cross-sectional data with the longitudinal data requiring an additional processing step. The longitudinal data processing must account for multiple scans of the same individual at various time points, and so requires within-subject registration step or T1 averaging step (i.e., the creation of a within-subject average T1 template creation). Subjects that had only one scan over all time points did not require this within-subject registration step.

To account for the multiple time points of the same subjects in longitudinal data, average T1 images were calculated for every subject with 2 or more scans. Average T1 images were calculated using SPM12 Serial Longitudinal Registration with default parameters. To avoid any bias introduced by averaging the multiple scan subjects, I also created a 'pseudo-average' for participants with only a single scan. To do this, T1 images were flipped left-to-right; the original and flipped images were

then registered and an average image created (Figure 4.1). The longitudinal registration produced deformation fields that contain parameters to warp each time point T1 to the subject-specific T1. These parameters can then be used to warp the associated BIANCA-produced WMH maps into T1 average space, prior to normalisation.

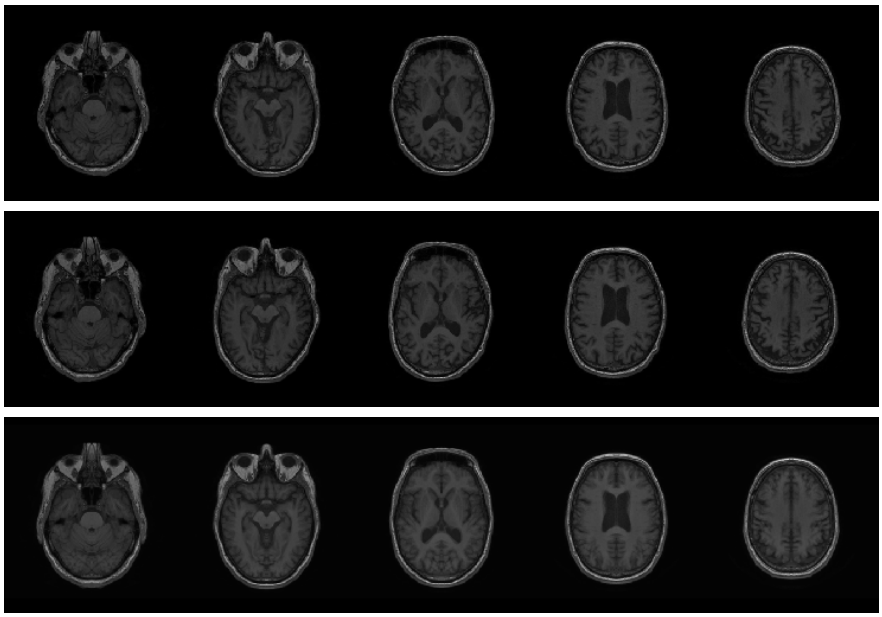


Figure 4.1: Example of a single scan subject (a) original T1 image, (b) left-to-right flipped image, and (c) averaged image.

4.6 Segmentation and normalisation

4.6.1 Cross-sectional

CAT12 (r1278, http://www.neuro.uni-jena.de/cat/), a toolbox of SPM12 (v7219, http://www.fil.ion.ucl.ac.uk/spm/), running in Matlab 9.3.0.713579 (R2017b), was used to process T1-weighted structural images. Briefly, images were bias-corrected, spatially normalised via DARTEL (using the DARTEL template provided within CAT12, registered to MNI space), modulated to compensate for the effect of spatial normalisation, and classified into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF), all within the same generative model [3]. The segmentation procedure was extended by accounting for partial volume effects [55], applying adaptive maximum a posteriori estimations [40], and using a hidden Markov Random Field model [11]. Coregistared WMH lesion masks in T1 space were warped into MNI space using the structural

DARTEL parameters.

4.6.2 Longitudinal

This step is identical to the cross-sectional segmentation and normalisation, however, the within-subject template image (i.e., average T1 image) is segmented and normalised. The deformation fields encompassing the warps from individual time point subjects space to average T1 space, and average T1 space to standard space (DARTEL) where then combined into a single deformation field to take individual BIANCA output mask from T1 subject space to standard space in one step. The deformation composition was done using the SPM12 tool Deformations. This tool combines the two separate deformation fields into one. In the same step as combining the two deformation fields, the application of the new deformation field can be applied to the BIANCA output mask in T1 space to warp the mask in standard space. The resultant image at the end of both the cross-sectional and longitudinal steams is a normalise (MNI space) WMH map per person per time point.

4.7 Smoothing BIANCA output mask in standard space

After normalising, each BIANCA output mask was smoothed using an fwhm= $8mm$ Gaussian kernel in normalised space. Smoothing was an important step before statistical analysis because smoothed masks adjust for potential registration errors or inconsistencies in the normalisation process. Voxel-wise statistical analyses were performed on the smoothed, normalised WMH masks and the specific statistical tests used in this thesis are outlined in Section (4.10).

4.8 Regional white matter hyperintensity definition

In addition to voxel-wise comparisons, I was interested in investigating the relationship between regional WMHs and cognition in PD. I extracted lobar WMH volume, as well as periventricular WMH volume from each participant at each time point from the normalised WMH mask. Brain lobes (frontal, occipital, temporal, and parietal) were defined using the Montreal Neurological In-

stitute (MNI) lobe atlases (available with FSL, Figure 4.2b).

The periventricular region was defined by dilating the FSL-provided ventricle mask in standard space (MNI space) using a spherical kernel with diameter = $7mm$ [18]. This periventricular region was then segmented further into anterior and posterior regions by using the frontal lobe junction to the other three brain lobes as a separation point (Figure 4.2c).

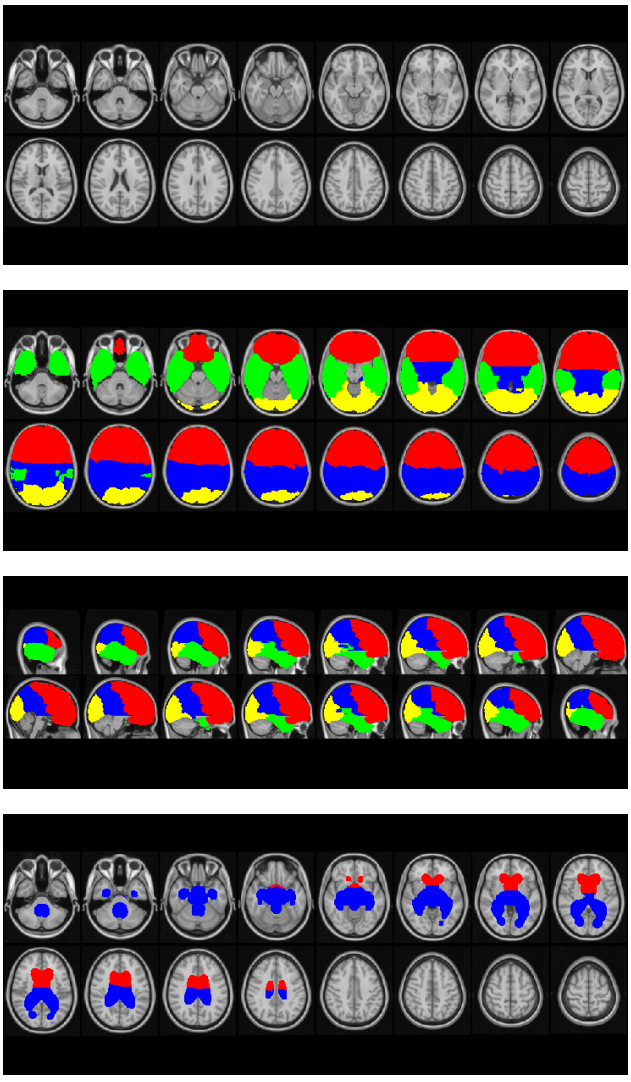
In addition to global volume, the following regional volumes were investigated in this thesis; frontal lobe, parietal lobe, temporal lobe, occipital lobe, total periventricular region, anterior periventricular region, and posterior periventricular region.

4.9 Analysis: Lesion probability maps

Lesion probability maps (LPMs) are voxel-wise representations of the likelihood of a single voxel being classed as a WMH. Group-wise LPMs were calculated at baseline in all cognitive categories (Control, PD-N, PD-MCI, and PDD) by averaging the normalised, smoothed baseline BIANCA WMH masks. LPMs are a way of visually indicating of the distribution of WMH burden by group.

4.10 Analysis: Statistical methods

I analysed BIANCA output masks using two statistical approaches to determine the potential importance of WMH burden on cognitive decline in PD. The General Linear Model was used to assess the difference in the spatial variation of WMHs between different cognitive groups on a voxel-wise level, as well as any voxel-wise association between global cognitive ability (as a continuous measure) and WMH volume. When investigating global or lobar WMH volume (i.e., as a single volume per region), I used Bayesian regression modelling to assess the out-of-sample predictive power of WMH burden to predict cognitive function. These two statistical analysis methods are detailed in the following section.



(a)

(b)

(c)

(d)

Figure 4.2: (a) Standard MNI152 (Montreal Neurological Institute) T1 1x1x1mm brain image. (b) Coronal view of brain lobes; frontal (red), occipital (yellow), temporal (green), and parietal (blue). (c) Sagittal view of brain lobes; frontal (red), occipital (yellow), temporal (green), and parietal (blue). (d) Coronal view of periventricular region; anterior (red) and posterior (blue).

4.10.1 General linear model

The general linear model (GLM) is an analytical method by which prediction of a dependent variable is carried out with a linear combination of independent variables. The GLM includes statistical techniques such as Student's t-test, multiple linear regression, analysis of variance (ANOVA), and analysis of covariance (ANCOVA) [43]. The GLM includes allowances for predictive variables, and it functions as a method of evaluating the relationship between constituent variables to explain the variability in a response variable [20].

The basic mathematical expression of the GLM is

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_q X_q + \varepsilon \tag{4.1}$$

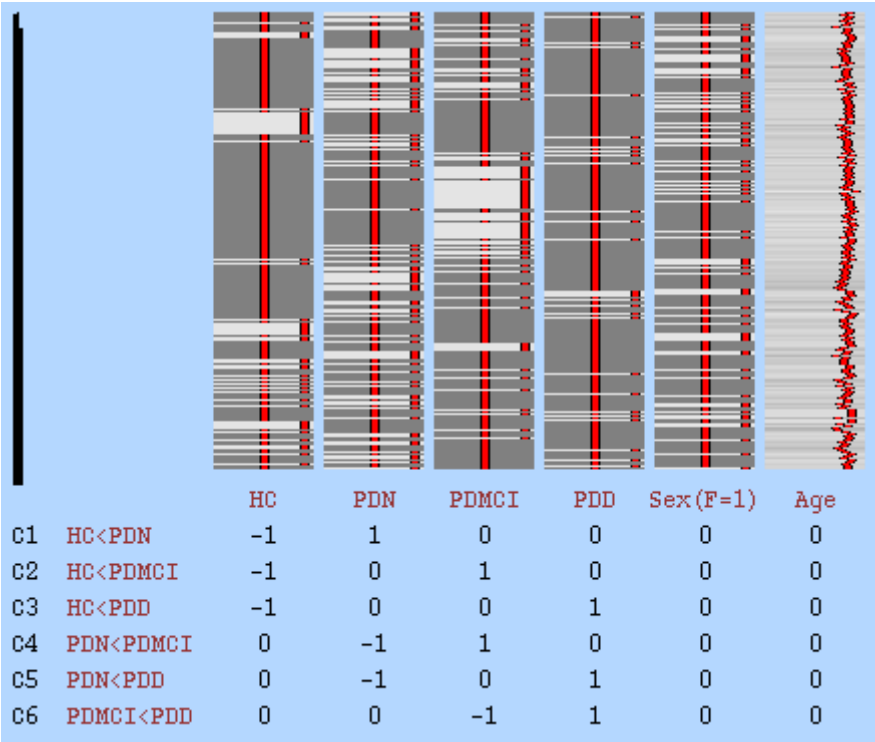
in which Y is vector response variable, $X_1, ..., X_q$ are the explanatory variables, $\beta_0, ..., \beta_q$ the regression coefficients, and ε is an error term to encompass any variation that cannot be described by the model [20]. In matrix notation, the GLM is written as follows in equation 4.2 where in Y, β, and ε are $n \times 1$ vectors for the $i = 1, 2, ..., n$ cases, and design matrix X is an $n \times q + 1$ matrix [43].

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1q} \\ 1 & X_{21} & X_{22} & \dots & X_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nq} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{4.2}$$

X is a matrix of the available data of the relationship under investigation, β is a vector of unknown and variable coefficients, and Y and ε are random vectors. The $q + 1$ rows in the X matrix of equation 4.2 account for the intercept term of the model, β_0. The $n \times 1$ error matrix is the random error variable and its constituents are assumed to be independently and identically distributed with a variance of σ^2 [43]. Another assumption utilised in the GLM is that the values of the dependent variable, Y, are a random selection of the population of interest.

4.10.2 GLM design matrix

Group differences (Control/PD-N/PD-MCI/PDD) in spatial WMH volume were assessed with age and sex as covariates. Results remained unchanged after inclusion of intracranial volume (ICV) as an additional covariate; results presented in this thesis are without ICV in the model and corrected for multiple comparisons using threshold-free cluster enhancement (TFCE) ($p < 0.05$). Figure 4.3 is an example of a design matrix used in this study computing the difference between cognitive groups WMH burden at baseline. Each different comparison of cognitive group is also known as a contrast of the GLM.



		HC	PDN	PDMCI	PDD	Sex(F=1)	Age
C1	HC<PDN	-1	1	0	0	0	0
C2	HC<PDMCI	-1	0	1	0	0	0
C3	HC<PDD	-1	0	0	1	0	0
C4	PDN<PDMCI	0	-1	1	0	0	0
C5	PDN<PDD	0	-1	0	1	0	0
C6	PDMCI<PDD	0	0	-1	1	0	0

Figure 4.3: GLM design matrix created using FSL Make GLM, identifying cognitive status as the tested effect and includes covariants sex and age at baseline. There are 6 separate pairwise comparisons tested, with tests specified by c1-6. Each column of the design matrix is binary, excluding age.
Legend: HC = Healthy Control; PDN = Parkinson's disease with normal cognition; PDMCI = Parkinson's disease with mild cognitive impairment; PDD = Parkinson's disease dementia.

In addition to the contrasts presented in Figure 4.3, a contrast between healthy controls and all PD patients at baseline was also tested to evaluate the significant difference between WMH burden in all controls and all PD patients. The contrast for this investigation is as follows (equation 4.3)

$$Y = \beta_{PD} - \beta_{HC} + 0\beta_{sex} + 0\beta_{age}. \tag{4.3}$$

4.10.3 FSL Randomise

The FSL function Randomise is employed to assess the contrasts set up in the GLM. Randomise is a statistical permutation method for statistical maps in which the null distribution is unknown and allows for the investigation of voxel-wise WMH load [58].

In this study, the inputs for Randomise include a 4D image file of smoothed, normalised BIANCA output files, the design matrix, and a binary mask excluding non-brain tissues. In this study, the number of permutations was set to 5000 for the building of the null distribution. All voxel-wise results were corrected for multiple comparisons using Family Wise Error Rate (TFCE ($p < 0.05$)).

The output from Randomise include a test statistic image and corrected p-value images for each contrast tested. Thresholds can then be applied to the p-value images to display only statistically significant differences in the test variables.

4.10.4 Bayesian regression model

The second statistical analysis method employed in this study was Bayesian multi-level regression models using the R package Bayesian Regression Models using 'Stan' (brms) (https:\\cran.r-project.org/web/packages/brms/brms.pdf). This method of analysis allows for cross-sectional analysis as well as determining the predictive power of WMH burden in relation to cognition over time. In each model, four chains with 4000 iterations each were used to generate posterior samples. Student-t priors were also included in the models and were set with 5 degrees of freedom, location = 0 and scale = 10, the default values.

brms analysis allowd for investigation into whether additional knowledge of WMH volume (global and regional) improves out-of-sample prediction of cognition in an individual when known predictors of cognition like age and PD diagnosis are already included in the model. brms analysis considers input variables, or predictors, of a model and calculates each variable's usefulness for predicting the outcome variable - cognitive score (global or domain) in this study. Individual predictors can be compared to determine their predictive power in any given model.

Similar to standard GLM-based regression, assessment of a particular predictor's usefulness in the Bayesian fit of the cognition model was evaluated, and considered significant if the upper and lower 95% confidence intervals (CI) for the variable fell exclusively above or below zero(that is, the effect of a particular predictor was considered significant if the 95% CI did not include 0).

While most studies stop with a description of the effects of parameters within a particular model, I took a further step by investigating if including a WMH volume variable provided any additional information about cognition compared to a simpler model. Model comparison was carried out using k-Fold cross-validation, a cross-validation technique in which the data is divided into k equal and random groups and $k-1$ groups are used to estimate a classifier (i.e. model prediction value), followed by the calculation of an error value by testing the calculated classifier on the group excluded from training [42]. A difference in k-Fold Information Criteria (KFOLDIC) score by at least twice the standard error of the estimated difference, indicated that there was a significant change in the out-of-sample prediction errors in any two compared models, where a lower KFOLDIC score represented an improvement in outcome variable (i.e. cognitive score).

4.10.5 Bayesian regression model comparisons

Question: Can global and/or regional WMH volume predict global cognitive z score in PD? *(Cross-sectional)*

Starting with an initial simple BRM of cognition modelled with just an intercept, predictors were added to the model one-by-one and assessment of the additional predictive power gained by the model from each variable addition was assessed using the KFOLDIC score. Predictors included age (in decades), group (control or PD), sex, and intracranial volume (ICV). The model with the best prediction of cognition using these input predictors was then used as a benchmark compared to models that also included WMH volume (global and regional) as an additional predictor. All 8 WMH regions specified above were compared to the benchmark model using k-Fold cross-validation. An example model is as follows: What is the contribution of global WMH load to the prediction of global cognitive score (cogz) when age, sex, intracranial volume, and group are already known:

$Cogz \sim age + sex + ICV + group + WMH(global)$ vs $Cogz \sim age + sex + ICV + group$.

Question: Can global and/or regional WMH volume predict specific cognitive domain scores in PD? *(Cross-sectional)*

A similar model comparison was also carried out using cognitive domain scores instead of global cognitive z score to assess if global or regional WMH volume could improve the predictive model of specific cognitive domain impairment. This was of particular interest given the dual syndrome hypothesis, detailed in Section 1.1.3, wherein WMHs in specific regions could have a more significant influence on the prediction of different cognitive domain scores. The cognitive scores used in the models were the mean values for the cognitive domains of interest; executive function, memory, visuospatial/visuoceptial, and attention. The regions used for this specific investigation included global, anterior periventricular, and posterior periventricular due to the specific investigation into the dual syndrome hypothesis in which the periventricular regions are the ones of most potential influence over cognitive scores in specific domains.

Question: Can baseline WMH volume be used to predict cognitive ability after 6 years? *(Longitudinal)*

The final models tested in this thesis used a longitudinal subset that included participants with both baseline and 6-year assessments ($Controls = 26$, $PD = 56$). Here, baseline variables were used to predict global cognitive z score at 6 years from baseline. The initial model used baseline age, baseline cognitive z score, and group (Control/PD) to predict global cognitive score at 6 years. I then added baseline global WMH volume to the model. In order to determine whether baseline global WMH volume increased the out-of-sample prediction of *future* 6-year cognition, the two models were compared as previously described using KFOLDIC.

Chapter 5

Results

Here, I present both cross-sectional and longitudinal results of the BIANCA-derived, normalised WMH masks analysis in the context of cognitive decline in PD. Cross-sectional results presented include:

- The investigation of total WMH volume across cognitive categories,
- Lesion probability maps, which display the distribution of WMHs,
- Baseline comparisons of the relationship between the spatial distribution of WMHs and cognitive decline in PD,
- And Bayesian regression models to specifically investigate the ability of WMHs to predict cognition both globally and by cognitive domain.

Longitudinal analysis investigated whether baseline WMH volume could improve the prediction of cognitive score at 6 years after baseline.

All results presented are based on the WMHs identified using BIANCA with optimised parameters, which I established in Part 1 of this thesis, with each WMH mask normalised according to the process detailed in Section 4.5.

5.1 Baseline cognition and white matter hyperintensity results

5.1.1 ANCOVA model

To investigate total WMH volume across differing cognitive groups (Control/PD-N/PD-MCI/PDD), I used an analysis of covariance (ANCOVA) model, with age and sex as covariates. The overall ANCOVA model was significant ($R^2 = 0.22$, $F(5, 222) = 14$, and $p < 0.001$). This was driven by an association between WMH volume and age. Tukey HDS post hoc tests were used for pairwise comparisons across the cognitive categories (included in Appendix A.1) in which there were no significant pairwise results across any of the different cognitive categories.

Baseline WMH volume across the 4 cognitive categories is displayed in Figure 5.1. The median volume values for the Control and PD-N group was comparable at 11.9ml and 11.7ml, respectively. The median values then increase for PD-MCI and again for PDD to 15.3ml and 17.6ml, respectively, but as documented above, these were not significantly different.

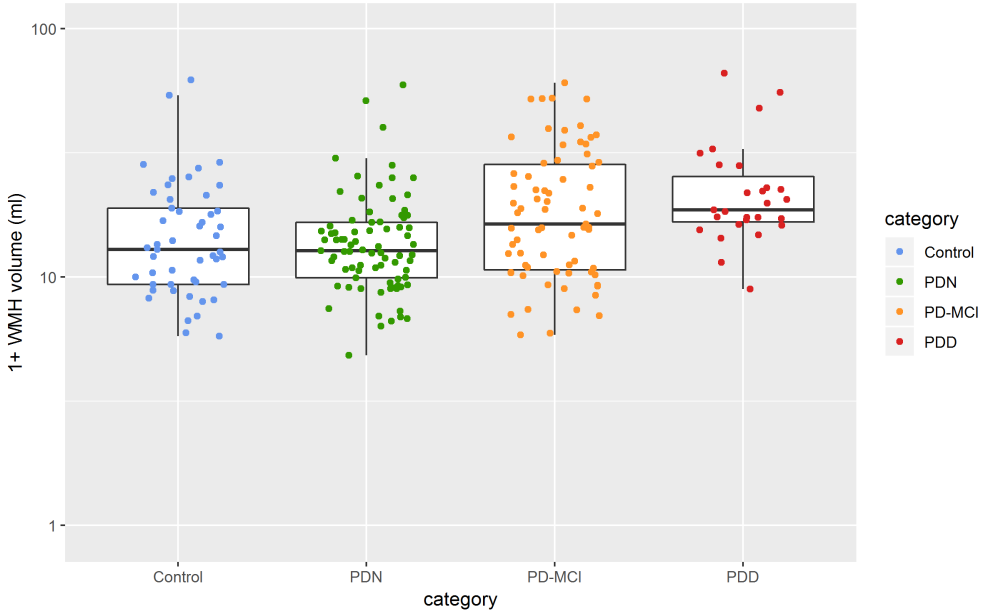


Figure 5.1: Baseline boxplot of white matter hyperintensity volume (WMH + 1) across categories; Control, PD-N, PD-MCI, and PDD. Each data point represents a unique subject with baseline MRI and neuropsychiatric assessment.
Legend: PDN = Parkinson's disease with normal cognition; PD-MCI = Parkinson's disease with mild cognitive impairment; PDD = Parkinson's disease dementia; WMH = White matter hyperintensity.

5.1.2 Lesion probability maps

Baseline LPMs were calculated for all cognitive categories (Control, PD-N, PD-MCI, and PDD), and visually show the extent of the likelihood of a region being hyperintense in the group. Even though global WMH volume did not differ statistically significantly by cognitive category, visually the LPMs become more extensive in PD as the cognitive status worsens (Figure 5.2) with PD-MCI and PDD LPMs growing slightly in size and probability as shown by the increasing yellow regions around the ventricles. The colour scale for these LPMs are all manually set corresponding to probability 0.05-0.5(5-50%) to emphasise the variability of the maps between the 4 categories as the maximum probability across all maps was 0.47(47%) in the PDD LPM (Figure 5.2d).

5.1.3 Randomise statistical comparison of spatial distribution across groups

While LPMs provide a purely visual representation of the distribution WMHs per cognitive category, I also statistically tested whether WMH distribution varied across these categories using a spatial GLM implemented in FSL. As for the global WMH ANCOVA model, here I investigated the relationship between spatial WMH volume and cognitive status, with age and sex as covariates. There were statistically significant areas WMH volumes in PD-MCI and in PDD relative to Controls (TFCE-corrected $p < 0.05$) as presented in the Figure 5.3. The significant regions appear predominantly on the right side of the brain in PD-MCI vs Control (Figure 5.3a), and the limited significant regions appear on the left side in PDD vs Control (Figure 5.3b). There were no other significant pairwise differences, i.e. there were no significant differences between PDN and Control, or within PDN, PD-MCI, and PDD. The lack of resounding pairwise comparison results from the Randomise comparison reflected the global ANCOVA results in Section 5.1.1.

I also restricted the analysis to PD only, in order to add disease duration, motor impairment (UPDRS part III), and LED to the model. There were no significant regions of difference between PDN, PD-MCI, and PDD (TFCE-corrected $p < 0.05$). This finding is again consistent with the Tukey statistical results that reported no significant results in the distribution of baseline WMH volume and cognitive category (Section 5.1.1).

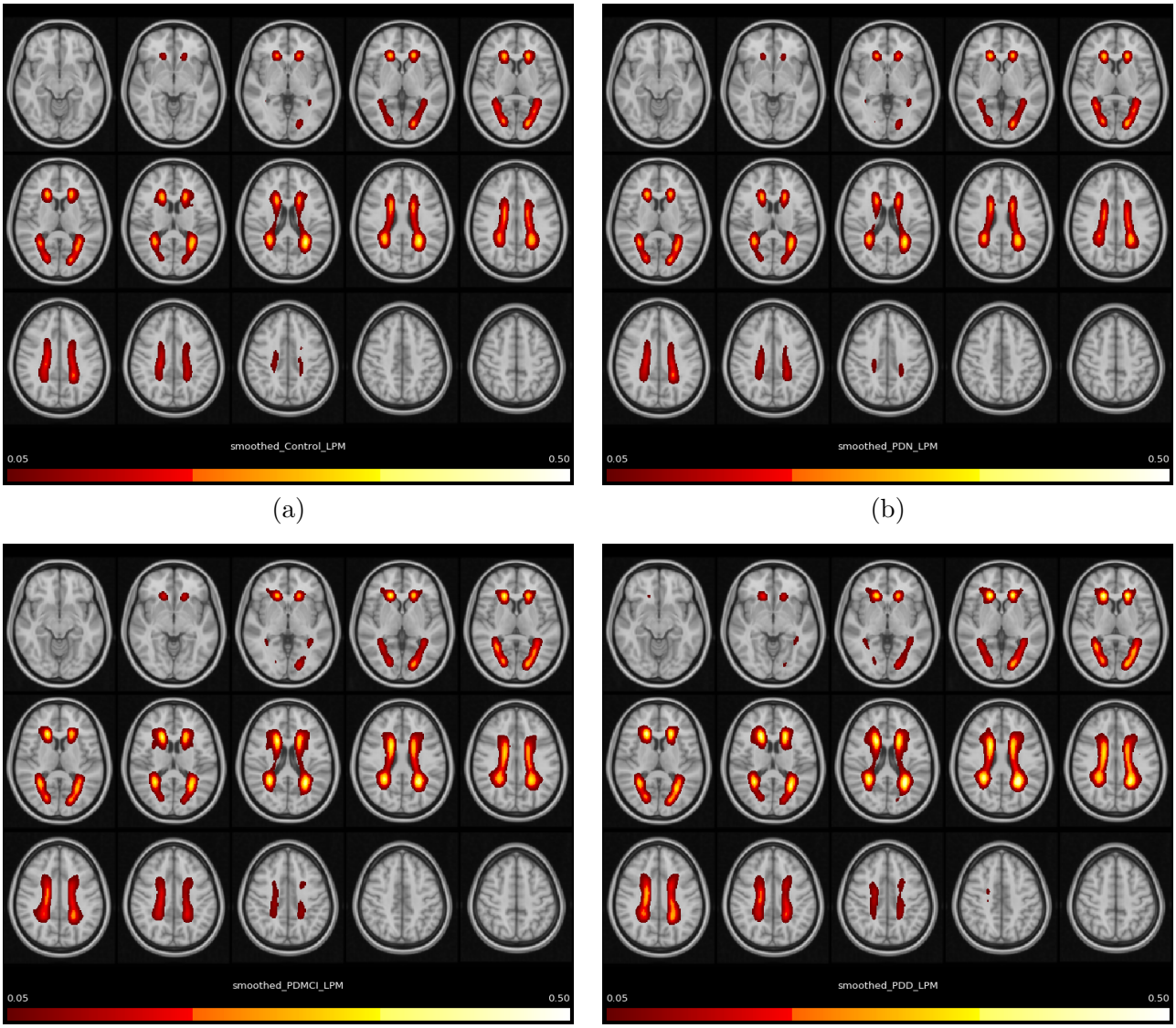
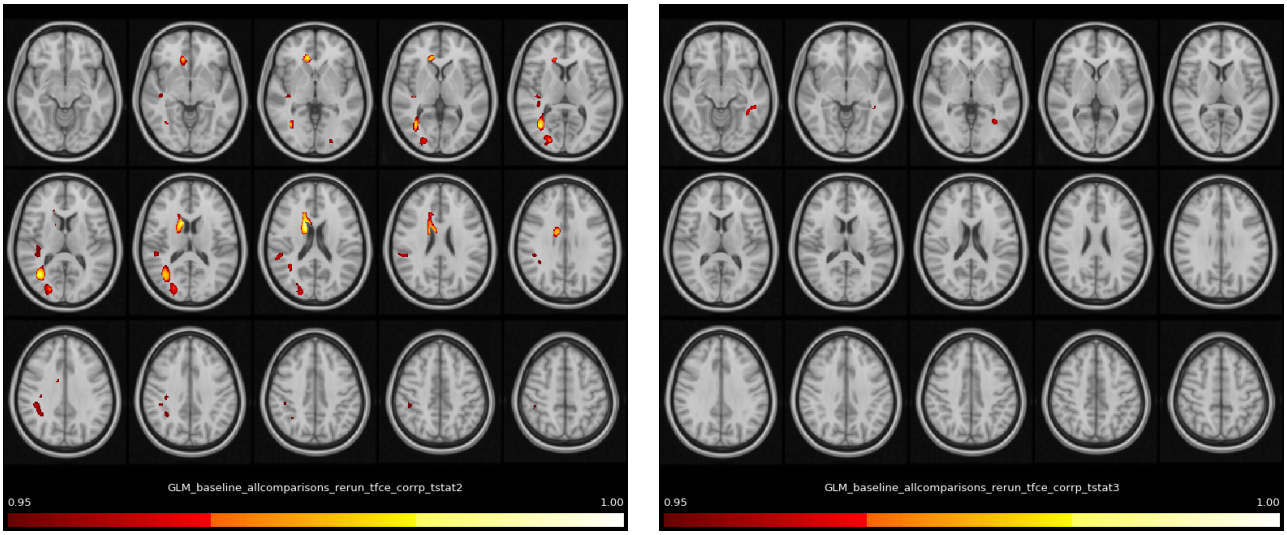


Figure 5.2: Lesion probability maps (LPMs) at baseline across the 4 cognitive groups. All data smoothed using a fwhm= 8 Gaussian kernel. Probability colour scale from 0.05 to 0.5. (a) Baseline Control LPM (b) Baseline PD-N LPM (c) Baseline PD-MCI LPM (d) Baseline PDD LPM.
Legend: PDN = Parkinson's disease with normal cognition; PD-MCI = Parkinson's disease with mild cognitive impairment; PDD = Parkinson's disease dementia.



(a) (b)

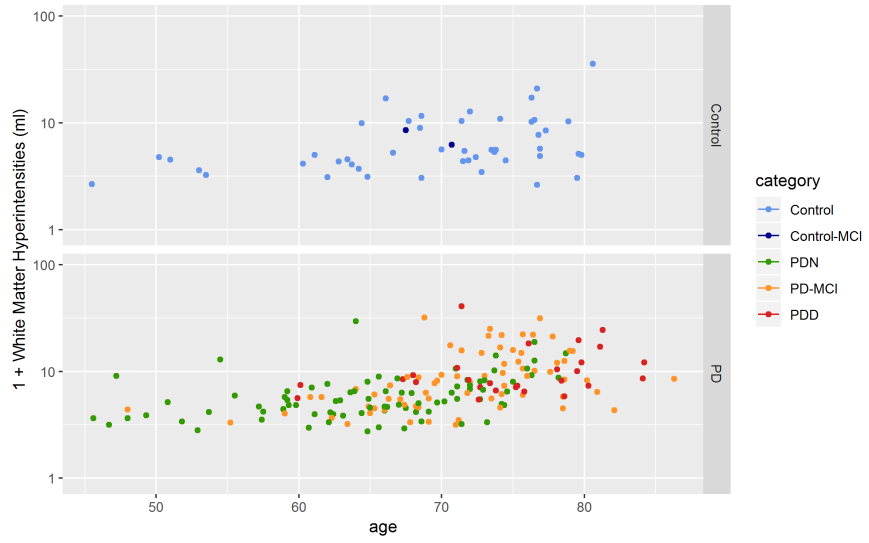
Figure 5.3: Red-yellow indicated significantly higher WMH volume at baseline in (a) PD-MCI vs Control and (b) PDD vs Control, accounting for age and sex. Significant results displayed for $p < 0.05$.
Legend: PD-MCI = Parkinson's disease with mild cognitive impairment; PDD = Parkinson's disease dementia.

5.2 Baseline white matter hyperintensity volume with age

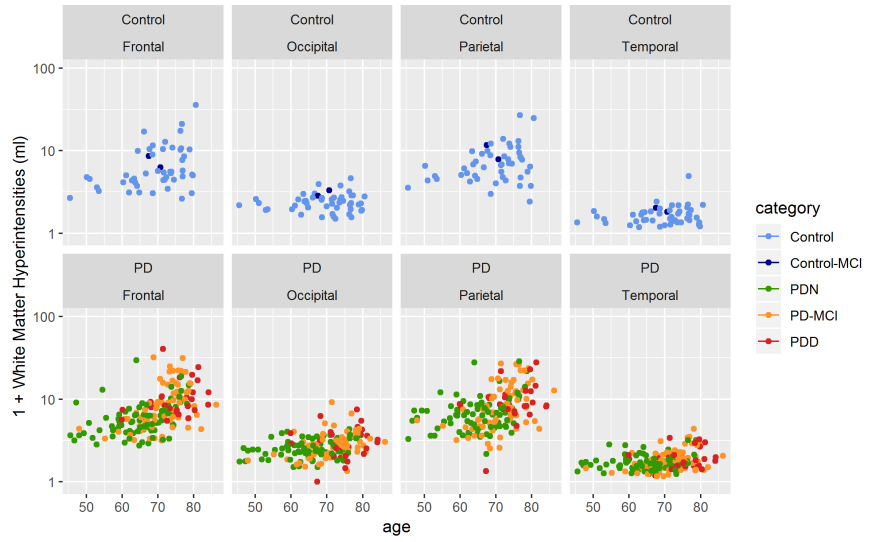
5.2.1 Global and regional white matter hyperintensity volume

Using a scatter plot at baseline, the relationship between increasing WMH volume and advancing age was investigated. Figure 5.4 shows WMH volume increased in older patients on both a global (Figure 5.4a) and regional level (Figure 5.4b). Note that there is a spread of cognitive status in which PD-N, PD-MCI, and PDD subjects have no distinctive separation of WMH volumes.

I also calculated correlation coefficients to quantify the relationship between global WMH volume and age. Table 5.1 presents the Pearson correlation coefficient values and accompanying p-values for each region presented in the age vs WMH volume scatter plot (Figure 5.4b), as well as additional periventricular regions. Correlation coefficients for total baseline data, both across the whole sample as well as within Controls and PD were statistically significant of $p < 0.01$. Limited statistically significants was calculated for correlation coefficients for exclusively baseline control subjects (Table 5.1).



(a)



(b)

Figure 5.4: Baseline WMH volume in control and PD (a) globally, and (b) regionally in lobes; Frontal, Occipital, Parietal, Temporal. The colour of points corresponds to the cognitive category of the subject at baseline as indicated by the key; Light blue = Control, Dark blue = Control-MCI, Green = PD-N, Yellow = PD-MCI, Red = PDD.
Legend: PD = Parkinson's disease; PDN = Parkinson's disease with normal cognition; PD-MCI = Parkinson's disease with mild cognitive impairment; PDD = Parkinson's disease dementia; WMH = White matter hyperintensity.

5.2.2 Randomise statistical comparison between age and white matter hyperintensity volume

Spatial correlation of WMH volume at baseline was investigated using Randomise to produce correlation maps with baseline age. Significant and extensive correlation was presented in the periventricular regions of the brain (Figure 5.5).

Region	PD and control correlation coefficient	Control correlation coefficient	PD correlation coefficient
Global	**0.46**(< 0.001)	**0.37**(0.010)	**0.49**(< 0.001)
Frontal	**0.50**(< 0.001)	**0.41**(0.003)	**0.52**(< 0.001)
Occipital	**0.28**(< 0.001)	0.07(0.626)	**0.33**(< 0.001)
Parietal	**0.40**(< 0.001)	0.31(0.030)	**0.43**(< 0.001)
Temporal	**0.25**(< 0.001)	0.19(0.182)	**0.28**(< 0.001)
Periventricular	**0.49**(< 0.001)	0.35(0.125)	**0.53**(< 0.001)
Anterior periventricular	**0.56**(< 0.001)	**0.45**(0.001)	**0.51**(< 0.001)
Posterior periventricular	**0.39**(< 0.001)	0.25(0.080)	**0.42**(< 0.001)

Table 5.1: Pearson correlation coefficients between regional WMH volume and age at baseline, with correlation coefficients calculated for all subjects (PD and control) and groups PD and control individually. Data presented as correlation coefficient(p-value). Statistically significant results are presented in bold.
Legend: WMH = white matter hyperintensity; PD = Parkinson's disease.

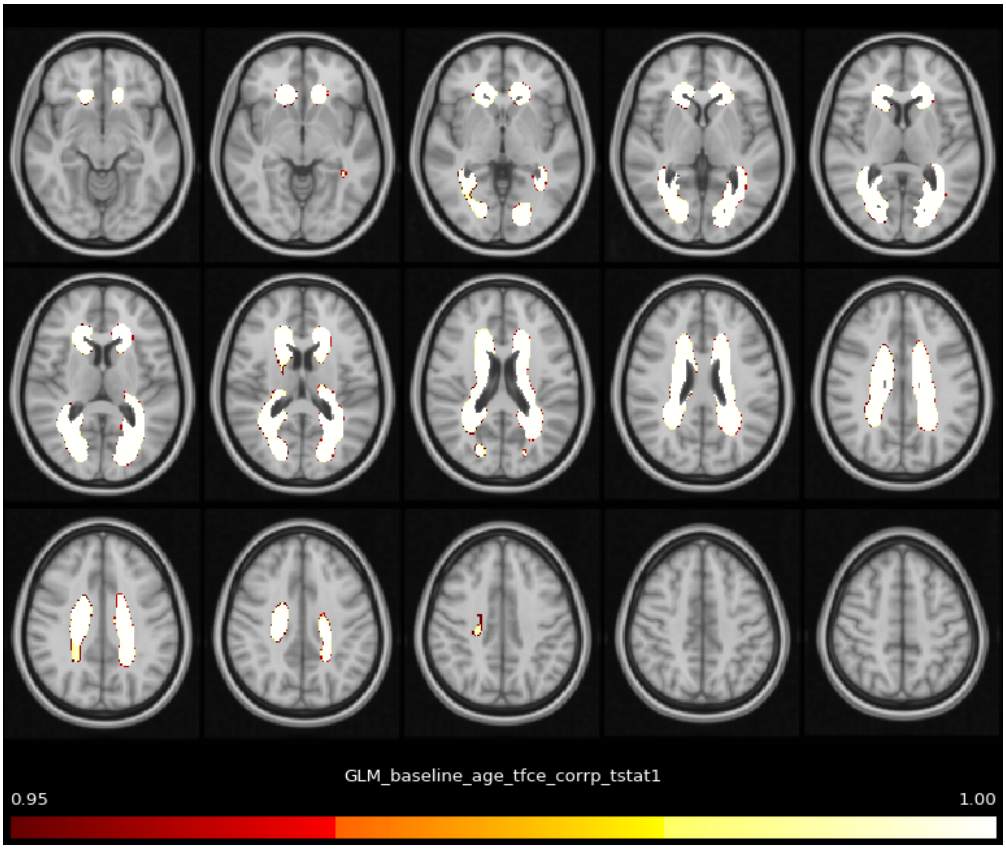


Figure 5.5: Baseline correlation map for white matter hyperintensity volume and age calculated at baseline. Significant results displayed for TFCE-corrected $p < 0.05$.

5.3 Cross-sectional Bayesian approach: Prediction of global cognitive ability

Cross-sectional analysis using the full cohort (i.e. all scans from all time-points were included and multiple scans per subject were modelled hierarchically) was carried out to investigate if WMH volume could be used as a predictor of current cognitive status, with and without additional data. An investigation into regional WMH effects on specific cognitive domains as well as global cognition was also carried out.

5.3.1 Global cognitive z score

Predictors not including WMH volume were compared to the initial model first (i.e. cognitive z modelled using only an intercept). Using KFOLDIC, it was determined that the inclusion of group (Control or PD) and age in conjunction improved the predictive power of the model significantly. The inclusion of sex and intracranial volume (ICV) did not improve the predictive power of the model further using the same k-Fold cross-validation, so they were excluded for all subsequent model comparisons.

From this point, models including group and age were then augmented to include WMH volume for all the individual regional WMH volumes under investigation, i.e. global, frontal, temporal, occipital, parietal, PV, anterior PV, and posterior PV regions; a new model was fitted for each region. None of the WMH regions investigated improved prediction of global cognitive z score when age and group were already included in the model (Table 5.2). All model comparisons are presented in Appendix A.2.

5.3.2 Cognitive domain scores

An investigation was also carried out to test the regional influence of WMH volume on specific cognitive domain scores. Attention, executive function, visuospatial, and memory and learning domains were tested using domain scores across all participants. The only regions of WMHs used for predicting cognitive domain scores were global, PV, anterior PV, and posterior PV since this investigation is specifically investigating the validity of the dual syndrome hypothesis in our cohort.

Model 1	Model 2	KFOLDIC	SE
Initial model	**Age + group model**	**104.09**	**22.50**
Age + group model	Age + group + global WMH volume model	-5.40	14.73
Age + group model	Age + group + frontal WMH volume model	-4.04	12.55
Age + group model	Age + group + temporal WMH volume model	-25.06	13.71
Age + group model	Age + group + occipital WMH volume model	-4.31	12.53
Age + group model	Age + group + parietal WMH volume model	-25.06	13.71
Age + group model	Age + group + PV WMH volume model	-6.56	12.41
Age + group model	Age + group + anterior PV WMH volume model	-5.58	14.29
Age + group model	Age + group + posterior PV WMH volume model	-18.90	10.86

Table 5.2: Summary of k-Fold cross-validation for models predicting global cognitive z score. Predictors included in the models were specified in each model. A positive KFOLDIC indicates an improvement between model 1 and model 2, with a KFOLDIC being considered statistically significant if it is at least double the value of the standard error. Bold text highlights models with significant improvement from models 1 and 2.
KFOLDIC = k-Fold information criteria; SE = standard error; PV = periventricular; WMH = white matter hyperintensity.

Cognitive domain	Model 1	Model 2	KFOLDIC	SE
Attention	**Initial model**	**Age + group model**	**84.54**	**21.06**
Attention	age + group model	age + group + global WMH model	-22.90	11.5
Attention	**Age + group model**	**Age + group + anterior PV WMH model**	**40.01**	**14.73**
Attention	Age + group model	Age + group + posterior PV WMH model	11.49	15.75
Visuospatial	**Initial model**	**Age + group model**	**41.16**	**15.41**
Visuospatial	Age + group model	Age + group + global WMH model	3.42	9.91
Visuospatial	Age + group model	Age + group + anterior PV WMH model	-11.98	9.39
Visuospatial	Age + group model	Age + group + posterior PV WMH model	0.84	8.11
Memory	**initial model**	**Age + group model**	**43.26**	**13.97**
Memory	Age + group model	Age + group + global WMH model	8.04	10.58
Memory	Age + group model	Age + group + anterior PV WMH model	14.74	11.02
Memory	Age + group model	Age + group + posterior PV WMH model	2.93	12.17
Executive function	**Initial model**	**Age + group model**	**56.80**	**20.25**
Executive function	Age + group model	Age + group + global WMH model	17.50	14.25
Executive function	**Age + group model**	**Age + group + anterior PV WMH model**	**41.27**	**17.35**
Executive function	Age + group model	Age + group + global posterior PV WMH model	-14.79	13.43

Table 5.3: Summary of k-Fold cross-validation information criteria values and standard error for models predicting mean cognitive domain score. Predictors included in the model comparison are specified in each model. A positive KFOLDIC indicates an improvement between model 1 and model 2, with a KFOLDIC being considered statistically significant if it is at least double the value of the standard error. Bold text highlights models with significant improvement from models 1 to model 2.
Legend: KFOLDIC = k-Fold information criteria; SE = standard error; PV = periventricular; WMH = white matter hyperintensity.

The domain scores KFOLDIC values were consistent to the global cognitive z score KFOLDIC value when comparing the initial model (intercept only) and the age + group model, with statistically significant improvements in KFOLDIC value in all tested cognitive domains (Table 5.3). There were also significant improvements reported in the models that included anterior PVWMH volume in the attention and executive function domain models. The comprehensive model comparison tables can be found in Appendix A.3.

5.4 Longitudinal white matter hyperintensity analysis

Longitudinal analysis was carried out in two ways; firstly using all available WMH volume and age data to assess longitudinally, and secondly using baseline and 6-year follow-up data to assess the power of baseline WMH volume in the prediction of future cognition.

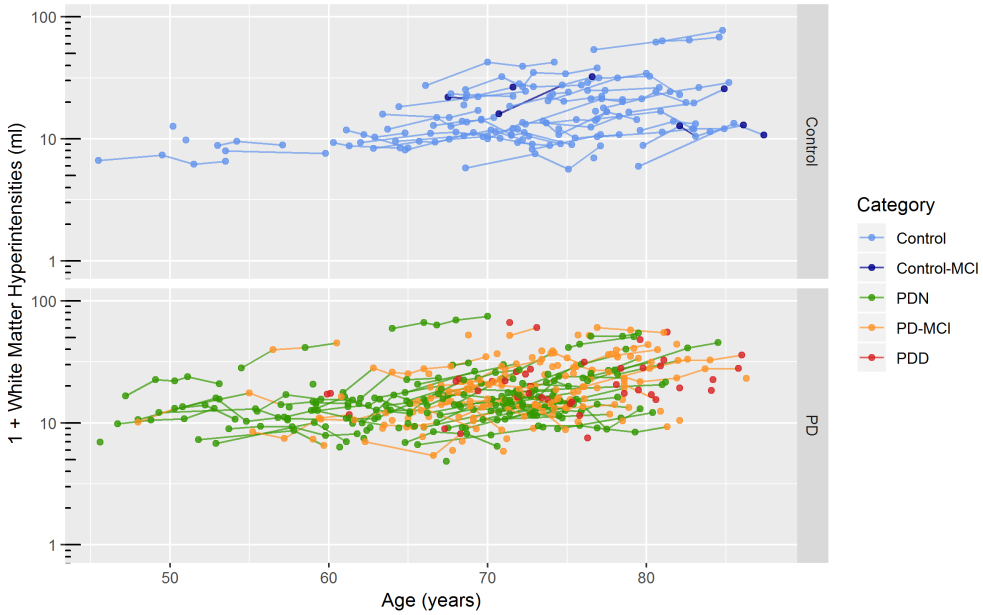


Figure 5.6: Global white matter hyperintensity volume at all collected time points. Each data point represents a MRI and cognitive assessment, and a line connecting two or more points joins follow-ups of an individual at multiple time points. Colour of each data point represents the cognitive status of the subject at the time of the assessment, and the colour of the line connecting two points corresponds to cognitive status at the first point.
Legend: PD = Parkinson's disase; PDN = Parkinson's disease with normal cognition; PD-MCI = Parkinson's disease with mild cognitive impairment; PDD = Parkinson's disease dementia.

5.4.1 Global white matter hyperintensity volume over time

Global WMH volume was correlated with advancing age when all data for all time points were considered (Figure 5.6), with Pearson correlation values for Controls $\rho = 0.44$, $p < 0.001$, PD $\rho = 0.41$, $p < 0.001$. It should be noted that high WMH volumes are also observed for PD patients with all cognitive statues (i.e. PD-N, PD-MCI, and PDD) as well as Controls, and the ranges of WMH volume are comparable across all cognitive categories .

5.4.2 Cognitive score 6 years from baseline

Figure 5.7 shows a plot of WMH volume vs change in cognitive z score at 6 years from baseline. The Pearson's correlation coefficient was $\rho = -0.18, p = 0.001$. There is also an emerging pattern in the plot which suggests that better cognitive ability at baseline is related to better future cognition.

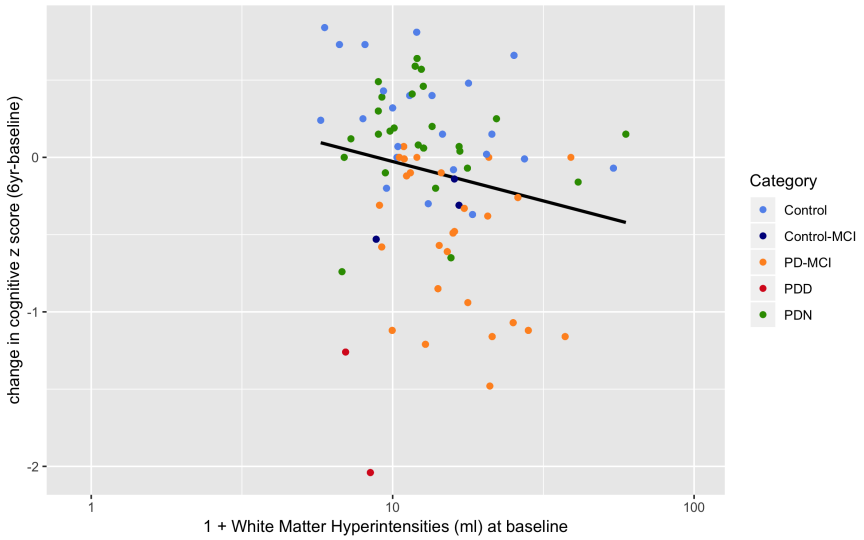


Figure 5.7: Baseline white matter hyperintensity volume vs cognitive z score change over 6 years. Cognitive category as indicated by data point colour is the category at 6 years from baseline. $\rho = -0.18$, $p = 0.001$.

5.4.3 Longitudinal model comparisons

Here I tested the predictive power of baseline parameters for cognition at a specified follow-up time point, 6 years from baseline (yfbl). This 6-year follow-up time restriction reduced the sample size to $n = 81$ for this particular analysis. Modelling cognition at 6 yfbl using baseline age, group (Control and PD), and baseline global WMH volume showed that global WMH volume improved the predictive model fit. The 95% confidence interval did not cross zero and thus was considered statistically significant. The effect size of the predictors of the 6 yfbl model are presented in Table 5.4a. The sign of the estimate value indicates the direction of change in cognition after 6 years. The positive estimate value for baseline cognitive z score indicates that a lower baseline cognitive z score is predictive of a lower 6 yfbl cognitive z score. The negative estimate values for baseline age and baseline WMH volume indicate increased age or WMH burden at baseline is indicative of a lower 6 yfbl cognitive z score. The positive estimate for group indicates that PD diagnosis at baseline is predictive of lower 6 yfbl cognitive z score.

Table 5.4b presents the model comparison of the 6 yfbl models, one without baseline WMH volume. The KFOLDIC value suggests there was no predictive value added to the 6-year cognition model when baseline WMH volume was included.

Predictors	Estimate	Estimate error	Lower 95 confidence interval%	Upper 95 confidence interval%
Intercept	**0.47**	**0.18**	**0.12**	**0.85**
Cognitive z score	**1.14**	**0.06**	**1.03**	**1.25**
Age (decade)	**-0.11**	**0.04**	**-0.19**	**-0.03**
Group (Control or PD)	**-0.43**	**0.07**	**-0.58**	**-0.29**
WMH volume	**-0.13**	**0.07**	**-0.26**	**-0.00**

(a) Summary of model results prediction models of cognitive z score at 6 years from baseline including WMH volume. All predictors are baseline values. All predictors had significant influence on future cognitive prediction by non zero crossing 95% confidence interval.

Model	KFOLDIC	SE
Baseline cognition, age, and group	489.54	29.77
Baseline cognition, age, group, and WMH	487.90	31.43
Model without WMH - Model with WMH	1.64	5.10

(b) KFOLDIC comparison between 6-year cognition prediction models. Significance difference in models shown a KFOLDIC value of two times the SE value.

Table 5.4: 6 years from baseline cognitive prediction model predictor estimates (a) and KFOLDIC values (b).
Legend: KFOLDIC = k-Fold inclusion criteria; SE = standard error; WMH = white matter hyperintensity.

Chapter 6

Discussion

6.1 Review of study results

The work done and presented in this thesis has been separated into 3 main investigations and a summary of the results are provided here.

Firstly, I determined that in our cohort, BIANCA produced WMH masks that most closely matched the gold-standard manual masks. BIANCA was further optimised by using bias-corrected input images and exclusion masks on the resultant BIANCA output WMH masks to reduce false positive WMH detection. I then applied BIANCA to the entire cohort to create WMH masks in all participants at all time points.

At cross-section, I showed a statistically significant relationship between increased WMH volume and advancing age. However, the analysis did not produce any statistically significant differences in WMH volume between cognitive categories in PD or compared to Controls. There was also limited significant results reported from the Bayesian model comparisons run at baseline for global and regional WMH volume in global z score and mean cognitive domain scores, with anterior PVWMH the only region that was found to improve the prediction of executive function and attention domain scores.

The longitudinal analysis utilised data from baseline and at 6 years follow-up. I found that al-

though baseline WMH volume was a significant predictor for future cognitive ability, it provided little additional information once age, group (Control or PD), and baseline cognitive ability were known.

In the following sections I will elaborate on the findings presented in Chapter 5 and summarised in here.

6.2 Baseline cognition and white matter hyperintensity volume

6.2.1 Global white matter hyperintensity volume

Baseline ANCOVA results suggested little evidence of an association between WMH burden and cognitive status. It is visually apparent in the boxplot (Figure 5.1) that high WMH volume is seen across Controls and the 3 constituent PD cognitive categories. The subtle median volume increase across the cognitive categories was very small, and the ranges of WMH volumes in all four categories were large, indicating substantial overlap in WMH volumes across all participants. Hence, the boxplot and associated ANCOVA model suggested that WMH volume did not differ across cognitive groups in PD. This in turn also indicated that WMH volume alone did not give a reliable prediction of the cognitive category on an individual level due to the overlap of WMH volumes presented across all categories.

6.2.2 Baseline lesion probability maps

The baseline LPMs make it visually apparent that the periventricular region is where WMHs are consistently likely to be located (Figure 5.2). With a number of cross-sectional studies reporting an association between PVWMH and worsening cognition in older adults [6], these LPMs support this finding, although not in a quantifiable, significant way. Visually, the LPMs did no appear to support the hypothesis that cognition worsens with increasing WMH burden.

6.2.3 Baseline spatial white matter hyperintensity distribution

Both PD-MCI and PDD exhibited significantly increased voxel-wise WMH volume relative to Controls, however, there were no significant differences in voxel-wise WMH volume across PD cognitive categories.

The PD-MCI vs Control comparison (Figure 5.3a) was more extensive than the PDD vs Controls comparison, however, there were no significant differences between PD-MCI and PDD when directly compared. The potential reason we may see the larger spatial extent of WMH increases in the PD-MCI vs Control comparison is the larger number of PD-MCI scans compared to PDD scans, and hence greater statistical power to identify a potentially small effect (71 PD-MCI vs 27 PDD). With this difference in category size in mind, it can be concluded that there may exist a slight increase in WMH volume anterior and posterior periventricular regions in PD-MCI and PDD compared to Controls.

6.3 Baseline age and white matter hyperintensity volume

I found evidence of the well-reported association between increasing age and increasing global WMH volume (Figure 5.4). This observation is in agreement with reviewed literature [24, 38, 53], and is observable both globally and regionally (Figure 5.5). Observation of this robust relationship acts as an internal check to processing methods used in this thesis. That is, while my primary question pertained to the relationship between WMH volume and cognition in PD, demonstrating the relationship between WMH volume and age in this cohort suggests that both imaging processing and analysis were not flawed or reporting misleading results. The analysis of the Pearson correlation coefficients in Controls and PD, both separately and as a single group supported the emerging association between age and WMH volume with statistically significant correlation coefficients reported for all brain regions in exclusively PDs and in Controls and PDs (Table 5.1). All these results collectively indicate that increased WMH volume, particularly periventricular, is correlated with advancing age in our cohort.

6.4 Cross-sectional Bayesian models

Cross-sectional model comparisons indicated that we once again see an effect of increasing age and increasing global WMH burden. There is, however, no evidence of increased WMH burden associated with PD or cognitive decline within PD, and a limited association between WMH burden and domain-specific cognitive ability (Table 5.2).

Interestingly, anterior PVWMH significantly increased out-of-sample prediction of attention and executive function domains (Table 5.3). Anterior dysfunction is implicated in worsening performance in both attentional processes and executive functions according to the dual syndrome hypothesis [27]. The results from this cross-validation investigation support this relationship with anterior PVWMH volume, significantly improving the appropriate models with the inclusion of anterior PVWMH volume. There is, however, a lack of support for the accompanying posterior dysfunction of the dual syndrome hypothesis which implicates worsening memory and visuospatial domain scores. The inconsistencies in the findings for the anterior and posterior PVWMH volumes and the influence of the volumes on the associated cognitive domains in our sample fails to provide convincing support or denial of the dual syndrome hypothesis.

6.5 Longitudinal analysis

6.5.1 Longitudinal age vs white matter hyperintensity volume

While WMH volume was shown to be related to increasing age, shown in this cohort and others, there were some subjects in this cohort that displayed decreasing WMH in time (Figure 5.6). While WMHs can become less intense over time in some situations (for example, in multiple sclerosis active lesions can change appearance on a T2 FLAIR image), this is relatively unlikely in an older, PD cohort. This suggested decrease in WMH volume in some subjects over time may represent inconsistencies in the WMH identification of BIANCA, even though BIANCA produced the best maps relative to the manually segmented gold-standard.

These inconsistencies in BIANCA could be because the algorithm doesn't include a longitudinal

pipeline, in which it would take into account the multiple scans from the same individual, which could improve the issue of apparently decreasing WMH burden over time. This considered, there is still an upward trend overall of WMH volume with increasing age as expected.

6.5.2 6-year follow-up cognition

The 6-year follow-up cognition investigation resulted in a weak correlation between increased baseline WMH volume and the amount of cognitive change experienced in the 6 years from baseline ($\rho = -0.18, p = 0.001$, Figure 5.7). This result suggests that baseline WMH volume could be used to predict future cognition, but volume could not be used alone to predict cognitive change. Perhaps additional variables such as age or baseline cognition could be used in conjunction with WMH volume to improve prediction, but at a subject level, baseline WMH volume could not be used to predict future cognition in PD or otherwise.

It should be noted that there are few PDDs included in this longitudinal analysis. This could be due to the loss of PDD participants across the 6-year follow-up period to discontinued participation in the study, or individual inability to complete MRI scan or cognitive testing. This, in turn, restricts the sample of PDs in this investigation to predominantly PD-N and PD-MCI participants. Having a data set more representative of all cognitive categories in PD could alter the predictive power of the baseline values employed by the model. However, the 6-year follow-up period restricts the data since individuals who digress from PD-MCI or even PD-N to PDD over the follow-up period may have difficulty completing the rigorous testing required at the follow-up time point.

6.5.3 6-year follow-up cognitive prediction models

6-year prediction models were tested to evaluate the ability of WMH volume to predict future global cognitive score. Using known baseline values to model known future cognitive scores at a consistent follow-up time point allowed me to assess if the inclusion of additional predictors in the models improved the prediction of future global cognition. At the group level, I showed a significant relationship between baseline global cognitive ability, age, whether a person had PD, and global WMH volume (Table 5.4a).

However, when investigating the predictive power of each variable on the individual level (assessed using KFOLDIC), baseline WMH volume did not add any additional information about 6-year cognitive ability (Table 5.4b). Baseline values of cognitive z score, age, and group did, however, provide independent and significant predictive information about future cognition. That is, all three variables provided additional, useful information about future cognitive ability. The apparent contradiction between the group-level results (WMH volume had a significant association with 6-year cognitive ability) and the individual-level results (once baseline cognitive z, age, and group status were known, WMH contributed no additional information) can be explained by the partial collinearity of the input variables. Baseline cognitive z, age, and group status contain similar information to global WMH volume; therefore, while a relationship exists on the group level, WMH volume fails to improve out-of-sample prediction of future cognition on the individual level. Not surprisingly, baseline cognitive z score provides the greatest predictive power for 6-year cognitive z score. That is, current cognition is the best predictor of future cognition (Table 5.4a).

6.6 Interpretation of results

The research carried out in this thesis has clinical implications as the relationship between cognition and WMHs in PD is divided in current literature, as discussed in Section 1.2.3. The results presented in this thesis indicate that WMH volume does not hold the significant predictive power suggested by some previous reports in this large, longitudinal cohort [16, 54]. Considering the results presented above, WMHs do not add any more information or power to cognition prediction models once other factors that have well established and supported correlations to cognition are considered, like age and PD diagnosis.

One possible reason WMH volume fails to provide convincing predictive power to cognition models is that the effects predicted by WMH volume are absorbed by confounding predictors used. The well established and supported correlation between cognition and age is an example of this, in which increasing age is correlated with decreasing cognition, specifically in PD patients [59]. In addition to this, I demonstrated there is a correlation between increasing WMH volume and age in

Section 5.4.1. With both increased WMH volume and worsening cognition clearly correlated with advancing age, there is the potential that the full predictive power of WMH volume on cognition is absorbed by the predictive power of age. The interconnectedness of these variables with each other could cause a 'watering down' of the prediction effects of WMH on cognition in PD. Age remains in the cognition prediction models because of the strength of predictive power it provides to the model and removing it to allow all the predictive power to go to WMH volume results in worse model fitting.

Clinically, the work done in this thesis indicates that WMHs should not be considered a clinically useful biomarker for cognitive decline in PD or even for the presence of PD. There are weak indications that specific WMHs located in anterior PV regions are more closely associated with attention and executive function, but the inconsistencies that arise from the posterior PV region adding no additional predictive power for visuospatial or memory domains remains unresolved.

6.7 Study strengths and limitations

The findings of this thesis generally align with much of the literature reviewed that reports no significant relationship between WMH volume and cognitive decline in PD. A major strength of this study is the large longitudinal cohort of PD patients and matched healthy controls used for comparison, providing the opportunity for both cross-sectional and longitudinal analysis of WMH volumes and cognition. Our cross-sectional cohort comprised 258 participants, a large cohort compared to most previous studies, with other PD specific cross-sectional cohorts of 111 participants (PD-N = 39, PD-MCI = 46, PDD = 26) [16], and 90 participants (PD-N = 65, PD-MCI = 25) [31] being reviewed.

Rigorous cognitive testing is a major strength of this study with all cognitive domains assessed and mean domain scores for each calculated that could be used for the regional effect of WMHs. Having a thorough cognitive battery as part of our study as well as consistent follow-up of participants is an added strength to this work.

In addition to having an extensive cross-sectional cohort, 81 of the 258 individuals in our cohort contributed to the longitudinal component of this data (Control = 26, PD-N = 26, PD-MCI = 27, PDD = 2). Having longitudinal cognitive and imaging data has added another layer to this study in which future cognition in PD could be investigated. Reviewed literature of a similar nature included a study by Dadar et al. (2018) that used a cohort of PD patients and healthy controls to test the significance of WMH volume in terms of cognition (Control = 174, PD = 365)[12]. We were unable to reproduce the results of the study by Dadar et al., which indicate that spatial information of WMH significantly improved prediction of cognition, and a higher WMH volume in PD patients is indicatived of faster cognitive decline compared to low WMH volume PD patients and high WMH volume matched controls. The discrepancies in these findings could be due to the different WMH identification methods and cognitive tests used in the respective studies, or the differences in cohorts, particularly the large number of Controls and PDs included in the study by Dadar and colleagues. A larger cohort could have made the correlation tests more sensitive to increases in WMH volume with worsening cognition, following the effect observed and discussed in Section 6.2.3.

The different statistical analysis method between this study and others should be considered as I extended the analysis beyond the group-level analysis where many other studies do not. The statistical test presented in Section 5.4.3 investigating 6-year follow-up cognition prediction reported similar findings at the group-level (Table 5.4a), when the statistical tests were investigated further to an individual level, the significance of the relationship between WMHs and cognition was not sustained (Table 5.4b). The inconsistency between group-level and individual-level statistical tests are due mainly to correlated variables used in the models, discussed in Section 6.5.3.

One limiting factor of this study is the potential overestimation of WMH volume by BIANCA in our cohort. Although a quarter of the BIANCA training subjects had low WMH burden specifically to address this issue, other low WMH burden subjects not included in the training data appeared to have false positives in the BIANCA output masks after post-processing masking and small cluster removal. As stated in the sampled literature, BIANCA has been reported to overestimated 'dirty WM', WM that bordered on hyperintense but have poorly defined margins [30]. The results from our study indicate higher mean WMH values compared to others in literature [12], with a mean

WMH volume value of 17 ml (range 4.4 - 73.6 ml). The larger mean value in comparison to other studies again emphasises the impact of overestimation of the low burden subjects seen in this study.

This overestimation was not completely eliminated from our cohort and is manifest in the smallest WMH volumes of $3.8cm^3$. Visual inspection of the subjects with the lowest WMH volume as calculated by BIANCA makes it clear that some dirty white matter has been picked up in low burden subjects, and are regions that could have been eliminated in manual WMH identification. However, the higher WMH volume subjects have consistently accurate WMH detection, and this overestimation for very low burden subjects is a trade-off considered in achieving the best estimation of WMH masks for the cohort as a whole. In addition, the accuracy of BIANCA across the high and low burden subjects in this cohort was found to be more consistent than the other algorithms tested, so the use of BIANCA for WMH identification remains justified.

Another potential limiting factor of this work could be present in the coregistration step. When coregistering the T2 FLAIR to the T1, I coregistered and resliced the WMH mask to T1 space, which then needed to be thresholded to return to a binary image. I later learned that this step could have been avoided by using the 'Coregister: Estimate' option in SPM. 'Coregister: Estimate' calculates the coregistration matrix between two images, and then adjusts only the header information of this image, without reslicing the image (which avoids interpolation inherent in reslicing). In the future, I would recommend avoiding unnecessary instances of reslicing data. However, all images in the cohort were processed in a consistent pipeline. Therefore, any noise generated by the extra reslicing step should be (1) minimal, and (2) consistent across all participants, which reduces the chance of affecting any of the statistical comparisons I investigated.

6.8 Future work

There is potential for future work following on from the findings presented in this thesis. One avenue of investigation could be into the significant improvement in cognitive score prediction in executive function and attention with the inclusion of anterior periventricular WMHs on the models in addition to age and group (PD and Control) (Section 5.3.2). Further investigation to this

finding could help to illuminate the implications this finding has on the dual syndrome hypothesis, and tend to the missing link of the posterior periventricular WMH relationship to cognitive domain scores we would expect to see if WMHs were a biomarker of the dual syndrome hypothesis.

There is also potential to further refine the WMH identification process. While BIANCA was clearly the best WMH identification algorithm in the training data set (initial set n=20 and extended set n=40), the issues of overestimation, as discussed above, could be improved. This would further strengthen the findings of this work. There is also potential for combining WMHs with other comorbidities or biomarkers to achieve a stronger relationship between WMH and cognitive decline in PD. This could include, but is not limited to, biomarkers present in other brain imaging modalities, neuropsychological tests or scores, or haematological biomarkers.

6.9 Concluding comments

The work carried out in this thesis consisted primarily of two main investigations; the identification of the optimal WMH identification algorithm to use on our large longitudinal cohort of PD patients, and to then use that algorithm's results to investigate the correlation between WMH volume and cognitive impairment in PD.

Of the four WMH identification algorithms investigated, BIANCA was determined to be the best for our cohort with the highest performance measures across the board. BIANCA was trained using 40 mixed WMH burden subjects, and bias-corrected input images to achieve the most accurate WMH identification. The resulting outputs produced by BIANCA were used to identify the correlation between WMH volume and cognitive impairment and worsening among PD patients, and compared to healthy controls.

Our baseline results supported the finding that WMH volume increases with age. The LPMs and the associated statistical tests, however, failed to convincingly support increasing WMH volume correlation with worsening cognition in PD in comparison to Controls and within the PD cohort.

Statistical maps calculated comparing PD-MCI and PDD patients with healthy controls produced restricted significant areas of increased WMH volume, notably more drastically for PD-MCI than PDD, possibly due to the larger number of patients in the cognitive group compared to PDDs allowing for more subtle effects to be detected. This finding loosely supports the claim that worsening cognition in PD is correlated to increased WMH volume when compared to healthy controls, but the same finding was not reported within PD. There is also no evidence from this investigation that there is any emerging spatial dependence of cognition on WMHs. Overall, these results provided weak support for increasing WMH volume and worsening cognition.

The cross-sectional predictive investigation of this study reported that specific regions of WMHs do not improve the prediction of global cognitive z score. When cognition is broken down further into cognitive domains, anterior PV regions emerge as improving variables for attention and executive function domain score prediction when age and group is already considered. These results align with one of the two syndromes in the dual syndrome hypothesis such that attention and executive function are suggested to be impacted more significantly by frontal cortical dysfunction, with anterior PVWMHs contributing to this dysfunction. However, lack of support for the other syndrome involving posterior cortical dysfunction effects on memory and visuospatial domain dysfunction renders this in need of further investigation.

For the longitudinal investigation in this study, baseline WMHs were used to test if they could be used to predict the future cognition of an individual. While group-level analysis reported a statistically significant estimate for WMH volume prediction of future cognition, subject-level investigation indicated no improvement in future cognitive function prediction. This suggests again that WMH volume does not add any predictive power of cognition when predictors such as age or PD diagnosis are already accounted for.

While WMH burden showed promise as a biomarker of cognitive decline in PD, this work has shown that they provide only a weak indication of future cognition, if at all. While the importance of WMH in ageing has been shown, WMHs do not appear the driving pathology of cognitive decline in PD. Other avenues of WMH biomarker detection will need to be investigated - potentially using

MRI/PET techniques. Perhaps WMHs in combination with other modalities will yield a better description of the underlying neural correlates of cognitive decline and dementia in PD, but this requires further investigation.

Bibliography

[1] D. Aarsland, B. Creese, M. Politis, K. R. Chaudhuri, D. H. Ffytche, D. Weintraub, and C. Ballard. Cognitive decline in Parkinson disease. *Nature Reviews. Neurology*, 13(4):217, 2017.

[2] T. J. Anderson. Genetics, brain imaging, and cognitive decline in parkinson's disease. Research Project Full Application (GA214F) to the Health Research Council of New Zealand., 2013.

[3] J. Ashburner and K. J. Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, 2005.

[4] K. Bendfeldt, P. Kuster, S. Traud, H. Egger, S. Winklhofer, N. Mueller-Lenke, Y. Naegelin, A. Gass, L. Kappos, P. M. Matthews, T. E. Nichols, E.-W. Radue, and S. J. Borgwardt. Association of regional gray matter volume loss and progression of white matter lesions in multiple sclerosis a longitudinal voxel-based morphometry study. *NeuroImage*, 45(1):60 – 67, 2009.

[5] N. I. Bohnen and R. L. Albin. White matter lesions in Parkinson disease. *Nature reviews. Neurology*, 7(4):229–236, 2011.

[6] N. Bolandzadeh, J. C. Davis, R. Tam, T. C. Handy, and T. Liu-Ambrose. The association between cognitive function and white matter lesion location in older adults: a systematic review. *BMC neurology*, 12(1):126–126, 2012.

[7] J. T. Bushberg. *The essential physics of medical imaging*. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 3rd edition, 2012;2011;.

[8] O. T. Carmichael, D. Drucker, and C. Schwarz. Longitudinal changes in cognition and cere-

brovascular disease in the Alzheimer's disease neuroimaging initiative. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 4(4):T277–T278, 2008.

[9] S. J. Chinta and J. K. Andersen. Dopaminergic neurons. *International Journal of Biochemistry and Cell Biology*, 37(5):942–946, 2005.

[10] B. S. Connolly and A. E. Lang. Pharmacological Treatment of Parkinson Disease: A Review. *JAMA*, 2014.

[11] M. B. Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, and J. . Thiran. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Transactions on Medical Imaging*, 24(12):1548–1565, 2005.

[12] M. Dadar, Y. Zeighami, Y. Yau, S.-M. Fereshtehnejad, J. Maranzano, R. B. Postuma, A. Dagher, and D. L. Collins. White matter hyperintensities are linked to future cognitive decline in de novo Parkinson's disease patients. *NeuroImage. Clinical*, 20:892–900, 2018.

[13] J. C. DalrympleAlford, L. Livingston, M. R. MacAskill, C. Graham, T. R. Melzer, R. J. Porter, R. Watts, and T. J. Anderson. Characterizing mild cognitive impairment in Parkinson's disease. *Movement Disorders*, 26(4):629–636, 2011.

[14] M. Emre, D. Aarsland, R. Brown, D. J. Burn, C. Duyckaerts, Y. Mizuno, G. A. Broe, J. Cummings, D. W. Dickson, S. Gauthier, J. Goldman, C. Goetz, A. Korczyn, A. Lees, R. Levy, I. Litvan, I. McKeith, W. Olanow, W. Poewe, N. Quinn, C. Sampaio, E. Tolosa, and B. Dubois. Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement Disorders*, 22(12):1689–1707, 2007.

[15] F. Fazekas, J. Chawluk, A. Alavi, H. Hurtig, and R. Zimmerman. MR signal abnormalities at 1.5 T in aAlzheimer's dementia and normal aging. *American Journal of Roentgenology*, 1987.

[16] R. Gonzlez-Redondo, J. Toledo, P. Clavero, I. Lamet, D. Garca-Garca, R. Garca-Eulate, P. Martnez-Lage, and M. C. Rodrguez-Oroz. The impact of silent vascular brain burden in cognitive impairment in Parkinson's disease. *European journal of neurology*, 19(8):1100–1107, 2012.

[17] A. M. Graybiel. The basal ganglia. *Current Biology*, 10(14):R509–R511, 2000.

[18] L. Griffanti, M. Jenkinson, S. Suri, E. Zsoldos, A. Mahmood, N. Filippini, C. E. Sexton, A. Topiwala, C. Allan, M. Kivimki, A. Singh-Manoux, K. P. Ebmeier, C. E. Mackay, and G. Zamboni. Classification and characterization of periventricular and deep white matter hyperintensities on MRI: A study in older adults. *NeuroImage*, 170:174–181, 2018.

[19] L. Griffanti, G. Zamboni, A. Khan, L. Li, G. Bonifacio, V. Sundaresan, U. G. Schulz, W. Kuker, M. Battaglini, P. M. Rothwell, and M. Jenkinson. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. *NeuroImage*, 141:191–205, 2016.

[20] R. F. Haase. *Multivariate general linear models*, volume no. 09-170;no. 09-170.;. Sage, Thousand Oaks, Calif, 2011.

[21] M. A. Hely, W. G. J. Reid, M. A. Adena, G. M. Halliday, and J. G. L. Morris. The sydney multicenter study of parkinson's disease: The inevitability of dementia at 20 years. *Movement Disorders*, 2008.

[22] I. Huertas, J. Silvia, F. Garca-Gmez, J. Lojo, I. Bernal-Bernal, M. Bonilla-Toribio, J. Martn-Rodriguez, D. Garca-Sols, P. Gmez-Garre, and P. Mir. Genetic factors influencing frontostriatal dysfunction and the development of dementia in Parkinson's disease. *PLoS One*, 12(4), 2017.

[23] J. Jiang, T. Liu, H. Liu, W. Zhu, R. Koncz, T. Lee, P. S. Sachdev, and W. Wen. UBO Detector A cluster-based, fully automated pipeline for extracting white matter hyperintensities. *NeuroImage*, 174:539–549, 2018.

[24] J. Jiang, M. Paradise, T. Liu, N. J. Armstrong, W. Zhu, N. A. Kochan, H. Brodaty, P. S. Sachdev, and W. Wen. The association of regional white matter lesions with cognition in a community-based cohort of older individuals. *NeuroImage: Clinical*, 19:14 – 21, 2018.

[25] J. Juntu, J. Sijbers, D. Van Dyck, and J. Gielen. Bias field correction for MRI images. In *Computer Recognition Systems*, pages 543–551. Springer, 2005.

[26] L. V. Kalia and A. E. Lang. Parkinson's disease. *Lancet (London, England)*, 386(9996):896, 2015.

[27] A. A. Kehagia, R. A. Barker, and T. W. Robbins. Cognitive impairment in Parkinson's disease: the dual syndrome hypothesis. *Neuro-degenerative diseases*, 11(2):79, 2013;2012;.

[28] J. H. Kim, K. J. Hwang, J.-H. Kim, Y. H. Lee, H. Y. Rhee, and K.-C. Park. Regional white matter hyperintensities in normal aging, single domain amnestic mild cognitive impairment, and mild Alzheimer's disease. *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia*, 18(8):1101, 2011.

[29] J. Langley, D. E. Huddleston, J. Sedlacik, K. Boelmans, and X. P. Hu. Parkinson's disease-related increase of T2*-weighted hypointensity in substantia nigra pars compacta. *Movement Disorders*, 32(3):441, 2017.

[30] Y. Ling, E. Jouvent, L. Cousyn, H. Chabriat, and F. De Guio. Validation and optimization of BIANCA for the segmentation of extensive white matter hyperintensities. *Neuroinformatics*, 16(2):269–281, Apr 2018.

[31] E. Mak, M. G. Dwyer, D. P. Ramasamy, W. L. Au, L. C. S. Tan, R. Zivadinov, and N. Kandiah. White Matter Hyperintensities and Mild Cognitive Impairment in Parkinson's Disease. *Journal of Neuroimaging*, 25(5):754–760, 2015.

[32] D. W. McRobbie, D. W. Moore, E. A.McRobbie, and M. J. Graves. *MRI from picture to proton (Third ed.)*. Cambridge, United Kingdom: Cambridge University Press, 2017.

[33] D. J. Myall, T. L. Pitcher, J. F. Pearson, J. C. Dalrymple-Alford, T. J. Anderson, and M. R. MacAskill. Parkinson's in the oldest old: Impact on estimates of future disease burden. *Parkinsonism & related disorders*, 42:78, 2017.

[34] J. A. Obeso, M. Stamelou, C. G. Goetz, W. Poewe, A. E. Lang, D. Weintraub, D. Burn, G. M. Halliday, E. Bezard, S. Przedborski, S. Lehericy, D. J. Brooks, J. C. Rothwell, M. Hallett, M. R. DeLong, C. Marras, C. M. Tanner, G. W. Ross, J. W. Langston, C. Klein, V. Bonifati, J. Jankovic, A. M. Lozano, G. Deuschl, H. Bergman, E. Tolosa, M. RodriguezViolante, S. Fahn, R. B. Postuma, D. Berg, K. Marek, D. G. Standaert, D. J. Surmeier, C. W. Olanow, J. H. Kordower, P. Calabresi, A. H. V. Schapira, and A. J. Stoessl. Past, present, and future of

Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy. *Movement Disorders*, 32(9):1264–1310, 2017.

[35] C. W. Olanow, F. Stocchi, and A. E. Lang. *Parkinson's disease: non-motor and non-dopaminergic features*. Wiley-Blackwell, Chichester, West Sussex, UK, 2011.

[36] L. Pantoni. Cerebral small vessel disease: From pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurology*, 2010.

[37] N. D. Prins and P. Scheltens. White matter hyperintensities, cognitive impairment and dementia: an update. *Nature Reviews Neurology*, 11(3):157–165, 2015.

[38] F. A. Provenzano, J. Muraskin, G. Tosto, A. Narkhede, B. T. Wasserman, E. Y. Griffith, V. A. Guzman, I. B. Meier, M. E. Zimmerman, A. M. Brickman, and A. D. N. Initiative. White matter hyperintensities and cerebral amyloidosis: necessary and sufficient for clinical expression of Alzheimer disease? *JAMA neurology*, 70(4):455, 2013.

[39] M. Rachmadi, M. del C. Valds-Hernndez, M. Agan, and T. Komura. Deep learning vs. conventional machine learning: pilot study of WMH segmentation in brain MRI with absence or mild vascular pathology. *Journal of Imaging*, 3:66, 12 2017.

[40] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport. Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, 16(2):176–186, 1997.

[41] W. Reginold, K. Sam, J. Poublanc, J. Fisher, A. Crawley, and D. J. Mikulis. Impact of white matter hyperintensities on surrounding white matter tracts. *Neuroradiology*, 60(9):933–944, 2018.

[42] J. D. Rodriguez, A. Perez, and J. A. Lozano. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 2010.

[43] N. J. Salkind, editor. *Encyclopedia of research design*. Thousand Oaks, CA: SAGE Publications, Inc., 2010.

[44] M. Saranathan, P. W. Worters, D. W. Rettmann, B. Winegar, and J. Becker. Physics for clinicians: Fluidattenuated inversion recovery (flair) and double inversion recovery (dir) imaging. *Journal of Magnetic Resonance Imaging*, 46(6):1590–1600, 2017.

[45] P. Scheltens, F. Barkhof, D. Leys, J. P. Pruvo, J. J. P. Nauta, P. Vermersch, M. Steinling, and J. Valk. A semiquantative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging. *Journal of the Neurological Sciences*, 114(1):7–12, 1993.

[46] P. Schmidt. *Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging*. PhD thesis, 2017.

[47] P. Schmidt, C. Gaser, M. Arsic, D. Buck, A. Frschler, A. Berthele, M. Hoshi, R. Ilg, V. J. Schmid, C. Zimmer, B. Hemmer, and M. Mhlau. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *Neuroimage*, 59(4):3774–3783, 2012.

[48] C. Seer, F. Lange, D. Georgiev, M. Jahanshahi, and B. Kopp. Event-related potentials and cognition in Parkinson's disease: An integrative review. *Neuroscience and biobehavioral reviews*, 71:691, 2016.

[49] Y. Shi and J. M. Wardlaw. Update on cerebral small vessel disease: a dynamic whole-brain disease. *Stroke and vascular neurology*, 1(3):83–92, 2016.

[50] E. E. Smith, D. H. Salat, J. Jeng, C. R. McCreary, B. Fischl, J. D. Schmahmann, B. C. Dickerson, A. Viswanathan, M. S. Albert, D. Blacker, and S. M. Greenberg. Correlations between MRI white matter lesion location and executive function and episodic memory. *Neurology*, 76(17):1492–1499, 2011.

[51] M. D. Steenwijk, P. J. W. Pouwels, M. Daams, J. W. van Dalen, M. W. A. Caan, E. Richard, F. Barkhof, and H. Vrenken. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). *NeuroImage: Clinical*, 3:462–469, 2013.

[52] P. A. J. Stoessl, P. S. Lehericy, and P. A. P. Strafella. Imaging insights into basal ganglia function, Parkinson's disease, and dystonia. *Lancet, The*, 384(9942):532–544, 2014.

[53] M. K. Sunwoo, S. Jeon, J. H. Ham, J. Y. Hong, J. E. Lee, J. . Lee, Y. H. Sohn, and P. H. Lee. The burden of white matter hyperintensities is a predictor of progressive mild cognitive

impairment in patients with Parkinson's disease. *European Journal of Neurology*, 21(6):922–e50, 2014.

[54] P. Svenningsson, Per, P. Westman, Eric, P. Ballard, Clive, and P. Aarsland, Dag. Cognitive impairment in patients with Parkinson's disease: diagnosis, biomarkers, and treatment. *Lancet Neurology, The*, 11(8):697–707, 2012.

[55] J. Tohka, A. Zijdenbos, and A. Evans. Fast and robust parameter estimation for statistiacl partial volume models in brain MRI. *NeuroImage*, 23(1):8, 2004.

[56] B. Vesel and I. Rektor. The contribution of white matter lesions (WML) to Parkinson's disease cognitive impairment symptoms: A critical review of the literature. *Parkinsonism and Related Disorders*, 22:S166–S170, 2015;2016;.

[57] L. O. Wahlund, F. Barkhof, F. Fazekas, L. Bronge, M. Augustin, M. Sjgren, A. Wallin, H. Ader, D. Leys, L. Pantoni, F. Pasquier, T. Erkinjuntti, P. Scheltens, and E. T. F. on Age-Related White Matter Changes. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke*, 32(6):1318–1322, 2001.

[58] A. M. Winkler, G. R. Ridgway, M. A. Webster, S. M. Smith, and T. E. Nichols. Permutation inference for the general linear model. *NeuroImage*, 92:381–397, 2014.

[59] K.-L. Wood, D. J. Myall, L. Livingston, T. R. Melzer, T. L. Pitcher, M. R. MacAskill, G. J. Geurtsen, T. J. Anderson, and J. C. Dalrymple-Alford. Different PD-MCI criteria and risk of dementia in Parkinson's disease: 4-year longitudinal study. *NPJ Parkinson's disease*, 2:15027, 2016.

[60] J. Young and M. Mendoza. Parkinson's disease: a treatment guide. *The Journal of family practice*, 67(5):276;279;284;286, 2018.

Appendix A

Cognition models

Table A.1: Tukey Post Hoc Test

Linear Hypotheses	Estimate	Standard Error	t value	Pr ($>|t|$)
PDN - Control == 0	0.075	0.088	0.854	0.825
PD-MCI - Control == 0	0.171	0.089	1.924	0.217
PDD - Control == 0	0.261	0.117	2.232	0.115
PD-MCI - PDN == 0	0.096	0.082	1.163	0.646
PDD - PDN == 0	0.186	0.115	1.623	0.363
PDD - PD-MCI == 0	0.090	0.108	0.833	0.836

PD-N = Parkinson's disease normal cognition; PD-MCI = Parkinson's disease mild cognitive impairment; PDD = Parkinson's disease dementia; MoCA = Montreal Cognitive Assessment.

Table A.2: Cognitive z score models

Cognitive z score models	KFOLDIC	SE
cogz.int.data	905.44	54.5
cogz.group.age.data	800.54	45.18
cogz.group.age.data	805.95	46.52
cogz.group.age.frontal.data	804.58	46.56
cogz.group.age.temporal.data	825.6	48.76
cogz.group.age.occipital.data	804.85	45.7
cogz.age.group.parietal.data	826.01	47.39
cogz.group.age.PV.data	807.1	46.09
cogz.group.age.antPV.data	806.12	46.18
cogz.group.age.postPV.data	819.45	46.74
cogz.int.data - cogz.group.age.data	104.9	22.5
cogz.int.data - cogz.group.age.data	99.49	21.9
cogz.int.data - cogz.group.age.frontal.data	100.86	20.57
cogz.int.data - cogz.group.age.temporal.data	79.84	17.44
cogz.int.data - cogz.group.age.occipital.data	100.59	20.49
cogz.int.data - cogz.age.group.parietal.data	79.43	20.86
cogz.int.data - cogz.group.age.PV.data	98.34	21.31
cogz.int.data - cogz.group.age.antPV.data	99.32	22.74
cogz.int.data - cogz.group.age.postPV.data	85.99	19.22
cogz.group.age.data - cogz.group.age.data	-5.4	14.73
cogz.group.age.data - cogz.group.age.frontal.data	-4.04	12.55
cogz.group.age.data - cogz.group.age.temporal.data	-25.06	13.71
cogz.group.age.data - cogz.group.age.occipital.data	-4.31	12.53
cogz.group.age.data - cogz.age.group.parietal.data	-25.47	14.18
cogz.group.age.data - cogz.group.age.PV.data	-6.56	12.41
cogz.group.age.data - cogz.group.age.antPV.data	-5.58	14.29
cogz.group.age.data - cogz.group.age.postPV.data	-18.9	10.86
cogz.group.age.data - cogz.group.age.frontal.data	1.36	12.35
cogz.group.age.data - cogz.group.age.temporal.data	-19.66	12.28
cogz.group.age.data - cogz.group.age.occipital.data	1.1	12.52
cogz.group.age.data - cogz.age.group.parietal.data	-20.07	14.84
cogz.group.age.data - cogz.group.age.PV.data	-1.15	12.33
cogz.group.age.data - cogz.group.age.antPV.data	-0.17	13.91
cogz.group.age.data - cogz.group.age.postPV.data	-13.5	11.63
cogz.group.age.frontal.data - cogz.group.age.temporal.data	-21.02	11.33
cogz.group.age.frontal.data - cogz.group.age.occipital.data	-0.27	10.99
cogz.group.age.frontal.data - cogz.age.group.parietal.data	-21.43	12.78
cogz.group.age.frontal.data - cogz.group.age.PV.data	-2.52	10.03
cogz.group.age.frontal.data - cogz.group.age.antPV.data	-1.54	10.62
cogz.group.age.frontal.data - cogz.group.age.postPV.data	-14.87	9.76
cogz.group.age.temporal.data - cogz.group.age.occipital.data	20.75	10.73
cogz.group.age.temporal.data - cogz.age.group.parietal.data	-0.41	11.58
cogz.group.age.temporal.data - cogz.group.age.PV.data	18.5	11.69
cogz.group.age.temporal.data - cogz.group.age.antPV.data	19.48	13.5
cogz.group.age.temporal.data - cogz.group.age.postPV.data	6.16	9.79
cogz.group.age.occipital.data - cogz.age.group.parietal.data	-21.16	11.49
cogz.group.age.occipital.data - cogz.group.age.PV.data	-2.25	10.84
cogz.group.age.occipital.data - cogz.group.age.antPV.data	-1.27	12.85
cogz.group.age.occipital.data - cogz.group.age.postPV.data	-14.6	9.05
cogz.age.group.parietal.data - cogz.group.age.PV.data	18.91	12.52
cogz.age.group.parietal.data - cogz.group.age.antPV.data	19.89	13.43
cogz.age.group.parietal.data - cogz.group.age.postPV.data	6.56	11.91
cogz.group.age.PV.data - cogz.group.age.antPV.data	0.98	10.29
cogz.group.age.PV.data - cogz.group.age.postPV.data	-12.35	9.57
cogz.group.age.antPV.data - cogz.group.age.postPV.data	-13.33	12.03

PD-N = Parkinson's disease normal cognition; PD-MCI = Parkinson's disease mild cognitive impairment; PDD = Parkinson's disease dementia; MoCA = Montreal Cognitive Assessment.

Table A.3: Cognitive domain models

Attention model kfold	KFOLDIC	SE
attn.int.data	961.71	42.74
attn.age.group.data	877.17	42.85
attn.age.group.wmh.data	900.07	45.02
attn.age.group.antPV.data	837.16	40.85
attn.age.group.postPV.data	865.68	41.75
attn.int.data - attn.age.group.data	84.54	21.06
attn.int.data - attn.age.group.wmh.data	61.64	23.06
attn.int.data - attn.age.group.antPV.data	124.55	25.16
attn.int.data - attn.age.group.postPV.data	96.03	18.17
attn.age.group.data - attn.age.group.wmh.data	-22.9	11.5
attn.age.group.data - attn.age.group.antPV.data	40.01	14.73
attn.age.group.data - attn.age.group.postPV.data	11.49	15.75
attn.age.group.wmh.data - attn.age.group.antPV.data	62.91	15.38
attn.age.group.wmh.data - attn.age.group.postPV.data	34.39	18.09
attn.age.group.antPV.data - attn.age.group.postPV.data	-28.52	18.71

(a) Attention domain model.

Visuospatial model kfold	KFOLDIC	SE
visuo.int.data	1104.36	46.3
visuo.age.group.data	1063.2	44.08
visuo.age.group.wmh.data	1059.77	45.16
visuo.age.group.antPV.data	1075.17	46.26
visuo.age.group.postPV.data	1062.36	44.63
visuo.int.data - visuo.age.group.data	41.16	15.41
visuo.int.data - visuo.age.group.wmh.data	44.59	15.7
visuo.int.data - visuo.age.group.antPV.data	29.19	16.14
visuo.int.data - visuo.age.group.postPV.data	42	14.82
visuo.age.group.data - visuo.age.group.wmh.data	3.42	9.91
visuo.age.group.data - visuo.age.group.antPV.data	-11.98	9.39
visuo.age.group.data - visuo.age.group.postPV.data	0.84	8.11
visuo.age.group.wmh.data - visuo.age.group.antPV.data	-15.4	9.55
visuo.age.group.wmh.data - visuo.age.group.postPV.data	-2.59	8.8
visuo.age.group.antPV.data - visuo.age.group.postPV.data	12.81	8.95

(b) Visuospatial domain model.

Memory model kfold	KFOLDIC	SE
memo.int.data	1411.53	36.67
memo.age.group.data	1368.28	36.86
memo.age.group.wmh.data	1360.24	36.2
memo.age.group.antPV.data	1353.54	36.88
memo.age.group.postPV.data	1365.34	36.05
memo.int.data - memo.age.group.data	43.26	13.97
memo.int.data - memo.age.group.wmh.data	51.3	14.94
memo.int.data - memo.age.group.antPV.data	58	15.19
memo.int.data - memo.age.group.postPV.data	46.19	15.38
memo.age.group.data - memo.age.group.wmh.data	8.04	10.58
memo.age.group.data - memo.age.group.antPV.data	14.74	11.02
memo.age.group.data - memo.age.group.postPV.data	2.93	12.17
memo.age.group.wmh.data - memo.age.group.antPV.data	6.7	9.77
memo.age.group.wmh.data - memo.age.group.postPV.data	-5.1	11.8
memo.age.group.antPV.data - memo.age.group.postPV.data	-11.8	10.89

(c) Memory and learning domain model.

Executive function model kfold	KFOLDIC	SE
exfn.int.data	1070.57	51.68
exfn.age.group.data	1013.77	48.98
exfn.age.group.wmh.data	996.27	48.39
exfn.age.group.antPV.data	972.5	44.63
exfn.age.group.postPV.data	1028.56	47.87
exfn.int.data - exfn.age.group.data	56.8	20.25
exfn.int.data - exfn.age.group.wmh.data	74.3	18.08
exfn.int.data - exfn.age.group.antPV.data	98.06	24.24
exfn.int.data - exfn.age.group.postPV.data	42.01	18.72
exfn.age.group.data - exfn.age.group.wmh.data	17.5	14.25
exfn.age.group.data - exfn.age.group.antPV.data	41.27	17.35
exfn.age.group.data - exfn.age.group.postPV.data	-14.79	13.43
exfn.age.group.wmh.data - exfn.age.group.antPV.data	23.77	17.8
exfn.age.group.wmh.data - exfn.age.group.postPV.data	-32.29	11.7
exfn.age.group.antPV.data - exfn.age.group.postPV.data	-56.05	16.36

(d) Executive function domain model.

PD-N = Parkinson's disease normal cognition; PD-MCI = Parkinson's disease mild cognitive impairment; PDD = Parkinson's disease dementia; MoCA = Montreal Cognitive Assessment.